EDITOR IN CHIEF

Dr Upali Rajapaksha

ISBN number: 978-955-0301-50-8

Published by
General Sir John Kotelawala Defence University
Ratmalana, 10390
Sri Lanka

Tel: +9411 337 0105
email: chair2018@kdu.ac.lk
website: www.kdu.ac.lk/irc2018

Published

ISBN number: 978-955-0301-50-8

PATRON, CONFERENCE STEERING COMMITTEE

Rear Admiral JJ Ranasinghe VSV USP psc MSc(DS) Mgt

PRESIDENT STEERING COMMITTEE

Major General IP Ranasinghe, RWP RSP ndu psc

CONFERENCE CHAIRMAN

Dr Upali Rajapaksha

CONFERENCE SECRETARY

Ms Bhagya Senaratne

ASSISTANT CONFERENCE SECRETARIES

Ms Nirupa Ranasinghe
Dr Danushi Gunasekara
Capt Madhura Rathnayake

STEERING COMMITTEE MEMBERS

President

Maj Gen IP Ranasinghe RWP RSP ndu psc

Brig RGU Rajapakshe RSP psc
Brig JMC Jayaweera psc
Capt (S) UG Jayalath
Commodore JU Gunaseela psc
Lt Col PSS Sanjeewa RSP psc
Lt Col WMNKD Bandara RWP RSP
Lt Col AK Peiris RSP
Capt MP Rathnayake

Professor MHJ Ariyarathne
Senior Professor JR Lucas
Senior Professor ND Warnasuriya
Senior Professor RN Pathirana
Senior Professor Amal Jayawardane
Dr (Mrs) WCDK Fernando
Dr KMG Prasanna Premadasa
Dr CC Jayasundara
Ms WB Senaratne
Dr (Ms) ADM Gunasekara

SESSION PRESIDENT

Commodore JU Gunaseela

PLENARY AND TECHNICAL SESSION COORDINATORS

Dr (Ms) W Gunathilake
Mrs DU Vidanagama
Mrs N Wedasinghe

ISBN number: 978-955-0301-50-8
PROCEEDINGS

SENIOR EDITORIAL BOARD

Mr. WAAK Amaratunga: President

Brig JMC Jayaweera psc
Commodore JU Gunaseela USP psc
Capt (E) KMD Senevirathne
Lt Col RA Samaranayke USP psc
Lt Col (Dr) MDAS Gunatilleke

Senior Professor WD Ratnasooriya
Senior Professor JR Lucas
Senior Professor ND Warnasuriya
Senior Professor (Mrs) SR De A Seneviratne
Senior Professor Amal Jayawandane
Senior Professor (Mrs.) BMR Fernandopulle
Senior Professor MHR Sheriff
Senior Professor ALS Mendis
Senior Professor RN Pathirana
Senior Professor TR Weerasooriya
Senior Professor (Mrs) N De Silva
Senior Professor PR Fernando
Professor SS Wickramasuriya
Professor W Abeywickrame
Professor J Welihinda
Professor E C S Vidyasekara
Professor (Mrs) IJ Dayawansa
Professor (Mrs) CL Goonasekara
Professor KN Pulasinghe
Dr (Mrs) RMNT Sirisoma
Dr (Mrs) LS Liyanage
Ms JADUN Jayasuriya
Ms DDL Willarachchi
Mr V Alagaratnam

EDITORIAL COMMITTEE

Prof K Pulasinghe
Dr LP Kalansooriya
Dr. (Mrs) HRWP Gunathilake
Mrs DU Vidanagama
Maj RMM Pradeep
Mrs N Wedasinghe
Mr PPNV Kumara
Mr B Hettige

ISBN number: 978-955-0301-50-8

FOREWORD

The International Research Conference 2018 of General Sir John Kotelawala Defence University (KDU IRC-2018) was on the 13th and 14th of September on the theme, Securing Professional Excellence through Collaboration. It was held for the 11th consecutive year under the guidance of the Vice Chancellor, Rear Admiral JJ Ranasinghe. The inaugural ceremony of the conference was held at the auditorium of the Faculty of Graduate Studies, under the patronage of the Secretary to the Ministry of Defence, Mr. Kapila Waidyaratne. Many distinguished guests: Tri-service Commanders, members of the Board of Management of KDU, representatives of the Diplomatic Corps, Vice Chancellors of other state universities, senior officers of the Tri-services and the Police, graced the event.

The session was opened by the Vice Chancellor, and he warmly welcomed all the dignitaries and participants. Further, he briefly explained the significance of the theme of the conference and KDU's commitment to fulfill national responsibilities. KDU IRC-2018 awarded honorary professorships to two internationally eminent Sri Lankan born scientists, Professor Mohan Munasinghe and Dr Sarath D Gunapala, in recognition of their contribution to their respective fields of science, and to mankind.

Delivering the keynote address, Mr. Waidyaratne commended KDU for playing a leading role in moulding the future of the military as well as civilian youth who are in pursuit of high quality tertiary education in Sri Lanka. He also stated that KDU contributed immensely to the much needed research and innovation, despite being an excellent institution for learning and disseminating knowledge that empowers the youth by helping them to develop sound attitudes and skills.

KDU IRC - 2018 continued with the tradition of bringing together researchers, academics and professionals from all over the world. This conference particularly encouraged the interaction of scholars to present and to discuss new and current research. Their contribution helped to make the conference as outstanding as it had been. A significant increase in the number of research papers received was noted at this conference. Out of 573 research papers received from both local and international scholars, 370 research papers were selected for presentation through the double blind peer review method. Each paper was reviewed by two independent experts in the field prior to selecting them for either oral or poster presentation. The selected papers were presented in nine research sessions, such as, Defence and Strategic studies, Basic and Applied Sciences, Engineering, Medicine, Allied Health Sciences, Computing, Built Environment and Spatial Sciences, Law and Management Social science and Humanities.

Technical Sessions were conducted on the first day of the conference in each faculty which drew approximately 55 guest speakers internationally and locally. Similarly, on the second day, parallel Plenary Sessions were conducted in the faculties under sub-themes, with the participation of approximately 370 experts delivering speeches related to their respective disciplines. The international guest speakers numbering more than 14 represented countries such as Japan, United States of America (US), United Kingdom (UK), India, New Zealand, Malaysia, Pakistan, Philippines, Burma, Indonesia, Bangladesh and Maldives,

KDU IRC-2018 was a unique research conference due to reasons, such as, international authors were facilitated to present via Skype remaining in their country; articles were automatically uploaded to Google Scholar in order to generate individual citations (H-indexing); the best papers of each category were published in the KDU Journal of Multidisciplinary Studies; and the best oral and poster presentation of each session were awarded.

This book contains proceedings of the sessions conducted under the disciplines of The plenary speeches and selected research papers presented at the technical sessions of the faculty are also included in this book, in addition to transcripts of the speeches delivered at the inaugural session. These Proceedings will no doubt furnish scholars of the world with an excellent reference book. I also trust that this will be an impetus to stimulate further study and research in all areas. I also trust that this would stimulate enthusiasm among scholars to engage in further study and to demonstrate the national and international importance of conducting research. I thank all authors, guest speakers and participants for their contributions.

A conference of this magnitude could not have been realized without the tremendous and generous support of the academic and administrative staff of KDU, who contributed to making it all happen.

Dr. Upali Rajapaksha
Editor
Conference Chairman 2018

ISBN number: 978-955-0301-50-8
PROCEEDINGS

ISBN number: 978-955-0301-50-8

Content

ISBN number: 978-955-0301-50-8

ISBN number: 978-955-0301-50-8
PROCEEDINGS

ISBN number: 978-955-0301-48-5

WELCOME ADDRESS

Rear Admiral JJ Ranasinghe VSV, USP, psc, MSc(DS) Mgt
Vice Chacellor

A very good morning to you!

I cordially welcome the Hon. Secretary to the Ministry of Defence, Mr. Kapila Waidyaratne, and I pay my gratitude to you Sir, for accepting our invitation and for being with us today at this 11th International Research Conference of General Sir John Kotelawala Defence University.

Next, I wish to extend a warm welcome to our Keynote Speaker, Prof. Mohan Munasinghe; and the Guest Speaker, Dr. Sarath D. Gunapala, both of whom are very eminent and distinguished Sri Lankan scholars who have made their imprint in the international arena. We are proud of your achievements and we consider your presence here today, as truly encouraging and inspiring us at KDU, as well as for all conference participants.

Let me also warmly welcome the Tri-service Commanders and all the other Members of the Board of Management of KDU. Also it is my pleasure to welcome Your Excellencies of the Diplomatic Corps; Vice Chancellors of other State Universities; and Senior officers of the Army, Navy, Air Force, and the Sri Lanka Police.

I also wish to extend a warm welcome to all dignitaries, scholars and participants; especially those of you from our friendly countries, who have come all the way to adorn this international conference in Sri Lanka.

Ladies and gentlemen, let me bid all of you present here today a very warm welcome this morning; and extend our appreciation for participating in this important event of our calendar.

We at KDU consider this annual conference very seriously due to several reasons. First, it is instrumental in establishing and strengthening the much needed research culture within the university, and it permeates the same into other universities and higher educational institutions in the country as well as into the industry through collaborations. Secondly, it gives local participants and institutions invaluable opportunities to establish links and networks with international counterparts, which is essential for progression in respective fields of specializations. Thirdly, it directly and indirectly contributes to the national growth and development in the long run. So, we consider this international research conference as an investment for the future.

As you are aware ours is primarily the National Defence University of Sri Lanka and our primary mandate is to produce academically and professionally qualified officers for our defence services, and we have been doing this for the highest satisfaction of the services. But today KDU has identified the need to establish firm civil military relations to face the complexities in national defence today, and hence the commencement of day-scholar programmes has helped us to achieve that goal while reducing the burden of the other state universities in providing adequate higher educational opportunities for our youths. The well-developed infrastructure, state-of-the art facilities as well as the dedicated human resources at KDU are now being meaningfully utilized to extend its services to deserving civilian youths to follow standard degree programmes in diverse disciplines, and the success of our achievement is seen in the increasingly higher number of foreign students who join our courses from countries, such as, the Maldives, India, Pakistan, Bhutan, Nepal, Uganda and Japan; along with expatriate students from Australia, Canada and the Middle East.

Ladies and gentlemen, in the modern times, it is essential not to compartmentalize varied fields, but to instill multidisciplinary collaboration among them. Hence relationships with different fields of innovation help to bridge gaps and inculcate professional excellence, which is the challenge of the 21st century. This explains the validity of the theme of our 11th International Research Conference, "Securing Professional Excellence through Collaboration".

KDU IRC is an ideal opportunity for the academia and professionals, to meet, discuss and exchange views in an academic environment. What is special about our conference is that, it is enriched with the participation of many local and foreign academics in varied disciplines; along with individuals from all three armed forces and the Police Department. Therefore, this is the only conference in Sri Lanka that brings together civilian professionals and their military counterparts.

I extend a warm invitation to the local and foreign students, academics and professionals present here today, to present their research findings; engage with other researchers in your field of study; have fruitful discussions and build life-long friendships with each other.

I welcome all to the 11th International Research Conference of KDU.

Have an inspiring and unforgettable day at KDU!

ISBN number: 978-955-0301-48-5

SPEECH OF THE CHIEF GUEST

Mr. Kapila Waidyaratne
President's Counsel Secretary

Good morning, everyone !

Chancellor, Vice Chancellor, Secretaries to the Ministries, Commander – Sri Lanka Navy, Chief of Staff of the Army, Dampath Fernando and the Air Force, Sumangala Dias, Your Excellencies of the Diplomatic Corp, international organizational scholars and other distinguished invitees.

It is my privilege and honour to be present here as the Chief Guest of the inauguration ceremony of the 11th International Research Conference organized by the Kotelawala Defence University, at which I happen to be the Chairman of the Board of Directors of its management. Let me first express my thanks to the Vice Chancellor, and conference organizers for inviting me as the Chief Guest of this very significant event.

Ladies and Gentlemen, as the Defence Secretary and also the Chairman of the Board of Management, I am aware of the outstanding role played by KDU in the tertiary landscape of Sri Lanka. With the donation of this beautiful estate along with the Kandawala mansion by the late General Sir John Lionel Kotelawala, the third Prime Minister of Sri Lanka, KDU was founded in the 1980s as the only tri-service academy in the country to provide much needed university education to the officers of the tri-services. Since then, KDU has come a long way over the last several decades reaching heights that may not have been dreamt at its inception.

For the last thirty years, KDU has produced thousands of graduate officers of very high calibre to lead the Army, the Navy and the Air Force, as highly disciplined and professional forces. In most recent times with its expansion to provide higher education opportunities to deserving civilian students, KDU has earned a name within and outside the country as a university that provides high quality tertiary education in diverse fields in a disciplined environment.

Today, with nine academic faculties, the Southern Campus and the recently established University Hospital, KDU has come to the forefront with determination to serve the nation in the best possible way. Therefore let me congratulate the Vice Chancellor and his able staff for the tremendous job, the excellent job done by them. Also let me take this opportunity to salute the pioneers of the university, specially the late General Sir John Lionel Kotelawala and Deshamanya Late General Dennis Perera, and let me not forget the political leadership of His Excellency the former President, J.R. Jayawardena, for the foresight to establish this University far back in the 1980s. Ladies and Gentlemen, the 11th International Research Conference that we are inaugurating today is a testimony for the significant role played by KDU in the field of higher education of Sri Lanka. As you are aware, it is not at all an easy task to successfully organize an annual conference of this magnitude considering the previous years. This itself indicates the strong commitment and responsibility of KDU to provide opportunities for the all-important task of knowledge creation and dissemination.

As you have already heard, and what I gathered from the Vice Chancellor, every year the number of research papers submitted for this conference is on the increase. It is heartening to know that thousands of researchers from all over the country as well as the world consider this conference an appropriate platform to present their papers. Therefore in my capacity as the Chairman of the Board of Management of KDU, I too share with KDU the pride of the leading role played by this defence university in popularizing research, which I believe is an essential, key aspect in the nation's growth.

Ladies and Gentlemen, the conference theme, Securing Professional Excellence through Collaboration, is timely particularly for countries like ours in our quest for appropriate development strategies in the face of new global challenges.

We do need meaningful collaborations across diverse professional bodies, and we cannot be completely looking after our own interests in isolated compartments. So time has come for all professionals to unite in sharing the burden of developing our nation economically, socially and culturally, so that the future generations will have a safer world to live in.
I believe it is our professional responsibility, irrespective of lables of distinction such as scientists, doctors, engineers, lawyers, academics, administrators, military professionals or any other, to find opportunities for innovative collaborations. And in such initiatives we all must reach excellence in our own professional domains and it is in this respect that universities and higher education institutions play an important role.

Ladies and Gentlemen, it is my belief that in this respect General Sir John Kotelawala Defence University is discharging its duty to the nation in a commendable manner. So let me conclude without taking much of your precious time as there are two eminent internationally recognized Sri Lankan intellectuals to deliver key note addresses at this conference. Let me once again thank the Vice Chancellor and the organizers for inviting me as the Chief Guest this morning. And let me also congratulate them for organizing a conference of this nature on a very timely and an important theme. Finally let me wish the two-day conference and both national and international participants a highly productive conference with intellectually stimulating deliberations.

Thank you very much.

ISBN number: 978-955-0301-48-5

KEYNOTE SPEECH

Professor Mohan Munasinghe
Professor of Sustainable Development, Sustainable
Consumption Institute, University of Manchester, UK
Founder Chairman of the Munasinghe Institute of Development, Sri Lanka

Good Morning to everybody!

Distinguished Audience, Vice Chancellor, Secretary, Service Commanders, Distinguished Guests Excellencies and of course fellow academics - I'm going to talk to you very briefly about Sri Lanka's sustainable mission and how we can achieve security, peace and prosperity through the green growth path. I would like to emphasize that we are looking for win-win solutions for people, the planet and prosperity for the entire globe. Now let me talk very briefly about the major issues that we face, i.e. threat for global security, and threats, such as, poverty and inequality due to resource shortages, shortfalls in the financial sector, disasters, conflicts and unfortunately weak leadership at the global level.

There is a concept called ecological foot print of humanity, which tells us how much of the planet resources we are using in total. In 2012, we were using one and a half times the equivalent of what the earth can sustainably produce, and by 2030 If we continue our present pattern of development we will need two planets! We know that we have only one planet. Sri Lanka is also exceeding its own ecological balance; it means we are chopping down the forests, polluting the water and so on. Now we have another question. It is the question of over consumption because if you look at who is doing the consumption, the richest people on the planet or the top 20% is consuming 85% of the resources, which is sixty times more than the poorest.

Sri Lanka has a dynamic nonaligned strategy; friend of all and enemy of none, which is something the President emphasized to me very much. The multipolar world order will be hopefully much more dependent on soft economic power rather than military power. And you have many poles of influence in the world. As we move to that, there will be disturbances, but Sri Lanka has a very key geostrategic position, and we can play a role in this. In case of climate change and global warming, there are two key facts which are the most important. The first point is; poor countries in poor groups suffer the most, which is manifestly unfair because the poor countries and the poorer people had the least to do with creating the problem. The problem was mainly created by the rich countries but the poor suffer. The second important point is that we follow this balance inclusive of a green growth path and make development more sustainable.

We can meet the challenge of climate change, as well as, all the other problems like poverty and so on. In the history, we have had many civilizations which lasted thousands of years. Whilst the Han civilization in China, Maurya Gupta Empires in India, Mesopotamian and the Roman Empire collapsed, eventually because of environmental and social factors, mainly over consumption of resources, there will be social divisions between rich elites and poor masses. So, these are very important aspects. We can learn a lot from the past history. If you take for example the hydraulic Systems in Sri Lanka, we had a wonderful

ISBN number: 978-955-0301-48-5

sustainable vision. For example, we believe that land belongs to the people and all living things while the ruler is only the guardian of the land; and King Parakramabahu had said not even a drop of rain water should flow into the ocean, without serving the man. If you look at the old dam anicuts, you will see that they were positioned exactly where the modern instruments tell us where they should be. They were ecofriendly and we had systems like the Velwidhana system and social system, controlling the flow of water which was extremely sustainable. So we have to be very much aware of these environmental and socio economic factors, scarcity of resources, inequality and conflicts which can also affect our present civilization. It could lead to some process of Barbarization where you have unrestrained market forces combining various problems like poverty, inequality and climate change, which would lead to a total breakdown of the planetary system. We also see the mass movement of refugees and other people which is more and more difficult to control, which is really a threat to global security in the future.

So, we now come to the last hope for mankind in a sense in this era, which is the 2015 sustainable development goals and the UN 2030 Agenda. How can we move forward towards a 21st Century Earth Eco-Civilization for a safer and better future? It is through the Balanced Inclusive Green Growth (BIGG) Path. The "Inclusive Green Growth Path", if you analyze the words- 'inclusive' means social; 'green' means Environment; and 'Growth' means Economy. These three elements are in the sustainable development triangle. And one of the core concepts that are extremely important here is, making development more sustainable. It is a call for empowerment and action. It basically says that sustainable development maybe very mysterious like a mountain peak covered with clouds. But we don't need to be discouraged. We will take one step at a time and climb up the hill, and eventually, we will reach the top. And the important thing here is that you and I, individuals, can make a difference.

We don't need to wait for Presidents, Prime Ministers and others to tell us what to do. Many of us know what we need to do. When we leave this room we switch off the light, we can turn off a tap, we can plant a tree, many things we can do that are extremely sustainable; so empowerment is extremely important. At the company level, we have corporate social responsibility and many other things. At the city level we can practice sustainable cities, and we come to the second core concept which is essentially what I told you before, that we need a prosperous economy specially with many millions of poor people in the world, we need to bring them out of poverty, but we also need to look at the environmental side that is the process of growth. So we don't want to destroy the environment and we need the social side which is the most neglected part. Unless we have social harmony none of the other things will be helpful. We can understand nature quite well, such as, forests, lakes and the air we breathe. But we have neglected social capital, human and cultural capital, and we had a 30 year war which eroded a lot of our social capital, this is the glue that binds the society together. All of us have a major role to play in that.

Just to remind you of the 2004 Tsunami in Sri Lanka, which is a shining example of social capital work. We were in the middle of a civil war, a poor country, one in every five hundred people was affected by the Tsunami, and other countries thought our society would collapse. But we rallied; people went out onto the beaches, voluntarily helped other people and cleared the bodies. There was social capital at work. If you look at the following year, in Hurricane Katrina in 2005 in one city New Orleans, what happened? There was no social capital: there was a complete breakdown of law and order, looting, raping and other things were going on. It was shocking because it's a very wealthy country and a small city. Hence social capital is not necessarily the property of the rich. Poor countries have effective social capital networks; we have it in Sri Lanka; we must build it and we must continue with it. I must also tell you very briefly that we need to transcend boundaries within our own mind; also this is for innovation that is what universities have to do.

Values are extremely important. We have to replace unsustainable values with more ethical values. We need to think in terms of multi-disciplinary issues. We need to think in terms of the whole planet and not just our own backyards. We have to think in terms of long time spans, decades and centuries; and as military practitioners, I'm sure you understand that it should be a long range plan, not just today or tomorrow. And we need to work with all stakeholders, i.e. the Government, not only the Government but also the civil society and businesses. Just to emphasize the question of social values, it is unethical social values that actually drive our society towards injustice, violence, greed and selfishness. That has created the state of what I call not as economic development, but as maldevelopment. We are growing based on debt, poverty, inequality and so on, which is not a very healthy way to go, and that has created what is called the environmental death over-using our planetary sources and also causing climate change. When we deplete our natural resources,

ISBN number: 978-955-0301-48-5

there is more conflict. So, you have unethical social values. That is a vicious cycle. If you look at the pattern of wars today, there are no wars on weapons of mass destruction, the wars are all for resources for oil, water and land.

So, this cycle has to be broken, and we have to transcend disciplines to do that, we need to think in multi-disciplinary terms, and we need to bring the civil and business societies to work with the government to push them to strengthen democratic space and provide good governance. So, let me just briefly tell you that climate change is in a precarious situation today. We have 280 parts per million, and the main indicator is the Co2 level in the atmosphere. That was at a safe level 100 years ago or more. Today it's over 400. So, we are exceeding the safe level of Co2, and what will happen is that we will have global warming, we will have more rainfall in wet areas so you have more floods, landslides; and more droughts in dry areas and more deserts; and we will have storms, cyclones and other things in addition to sea level rise and overall temperature increase. So the economic damage over the last 50 years is rising and it's going to continue to rise. We need to survive climate change, specially to protect the vulnerable people, poor children and the elderly, in some parts of the world, such as, small islands like the Maldives, and others which will be completely submerged, and particular sectors and systems like agriculture, coral reefs and so on, but unfortunately we are not doing those things.

Talking of sustainable production, there are two key points to consider. The first one is sustainability and triple bottom line, i.e. finance and economy, environmental and social. Those three have become much more important, it's not only a question of profits any more. The second one is effective usage of resources. If you are producing shoes, if you can produce shoes using less leather, less energy and less water; it is a win–win situation, because you are reducing the burden on the environment, and also reducing your cost. So this is very attractive and now we have technologies, which are win-win. There are many technologies which we have applied, for example, in case of a garment factory in Sri Lanka, MAS Holdings, which shows you how resource efficiency works. We have looked at how carbon and energy are used in the life cycle of the product from raw material to manufacturing, to distribution to use and disposal. The main carbon emissions of a garment come from raw material, not only from manufacturing. What is the lesson for us? If you want to reduce emissions of carbon, you are not going to tinker with the manufactory process; you are going to talk to your procurement officer. The procurement officer must buy raw material from the sources, which uses the least amount of carbon. It is not an engineering problem, it's a procurement problem. So this kind of analysis tells you where the hot spot is.

For energy, the hot spot is in manufacturing and distribution and in final endings. Why? Because people wash garments, that is energy, because of hot water. So these are methods on improving production processes; and let me just tell you that what we are planning here is to have sustainable consumers and sustainable producers working together because what you see in advertising today in the TV is mainly very unsustainable; it tells you to buy more and tells you to buy very unsustainable products. We have to break that side and eventually if we get these sustainable consumers and sustainable producers working together, we can eventually have a sustainable society and use modern tools. Traditional markets fall, so we can have organic markets and other things, where you go and buy the stuff, but for young people, it's online marketing. I'm supporting start-up companies which have huge sustainable online markets. Most young people who are in their twenties do not want to visit a shop; they go to the computer and shop online. So you have to have the tool to do the right thing and through these sustainable markets you can build a sustainable society.

In the sustainable Sri Lanka vision, we have hope for a thriving economy. We don't want to be poor. Being green and inclusive does not mean poor. We want to be prosperous and to lead a high quality life in an advanced stable economy, but it is green; it should be built on our traditional respect for nature, use resources efficiently and in an inclusive society. If you look into cross cutting issues, it has things like values, gender, international relations, security and peace; so all of these are integrated. This is one of the failures of Sri Lanka in every department of the Government. It is up to people like you, thought leaders, who can contribute to bring this integration about.

Let me just finish by saying we need to harmonize the economy, society and environment to build the democratic space in Sri Lanka. We need to work with the business society, civil society, the government, or all working together; and let me just end by reminding you that the situation in Sri Lanka has to be improved quite substantially because inequality has become much worse in the last two decades. Although GDP is growing, it is not reaching down. That is a very important aspect. Spatially also, the western province and so on are much better off than for example - the dry zone. We are not

ISBN number: 978-955-0301-48-5
PROCEEDINGS

investing enough in health; we are not investing enough in education rather low as a percentage of GDP.
There are also other things, for example, the Belt and Road Initiative (BRI). The Chinese government is very important for us strategically. If you want to be an Indian Ocean hub, we have to understand that we are right in the middle of the maritime BRR. So we need to be a friend of everybody and enemy of none. We have two major ports, Hambantota and Colombo right in the middle. So Sri Lanka's geostrategic position allows us to play a key role and the investments in the BRI will also help us to bring that about. But we have an important balancing act to play. And I think, the Foreign Ministry and security forces establishment of the country have a very important role in maintaining that balance and bringing prosperity to Sri Lanka. So, for the defense services you have to be good professionals; as professionals you have to be the best. But you have to also understand the economic, social and environmental dimensions of your job. And you have to broaden your perspective to bring those aspects this is difficult.

Although it is a difficult task, our graduates and others need to narrowly focus on their expertise and to be the best in the world. I think you can do much for building the nation, one nation and one flag, protecting the democratic space. And you have to understand the concept of National Identity. We all are Sri Lankans. We have a role in disaster, this is a peace time role going from conflict to resolution; through education and training, raising the standard of national conduct especially among young people; service to the nation; honesty; integrity; respect for nature and environment; respect for the society; law; tolerance and harmony; discipline; leadership; accountability; effectiveness and impartiality, and all of these values have to be rebuilt. My final message is that we face multiple problems, but we know how to address them. Unfortunately we need to do more; we have to go on the balanced inclusive green growth path. The Indian Ocean is a key area where we can do this. We need also to bottom up leadership; we don't have to wait for global leaders to tell us what to do. I think KDU and the Sri Lanka Defence Services can lead the way to peace and prosperity in the 21st century global civilization.

ISBN number: 978-955-0301-48-5

GUEST SPEECH

Dr Sarath D Gunapala
Solid-state Physicist and Senior Research
Scientist, Jet Propulsion Laboratory, NASA,USA

It's my pleasure to be here today, and I'm going to talk about the exploration of our solar system and beyond in the next thirty minutes or so. I have small stories to make it memorable, and I hope you will enjoy it.

I work at NASA Jet Propulsion Laboratory (JPL). Our main business at NASA is exploration of solar system and beyond, using robotic space-crafts. If you hear anything in the news, such as, going to Mars, Rovers, Jupiter, Saturn or beyond solar system, that is what we do.

We design and build space-craft, and seven minutes after launching we take control of it. With the space network, we can listen to our satellites even beyond the solar system. Two of our satellites, Voyager 1 and Voyager 2, are stationed about sixteen billion miles away, so if you have to send a radio signal at one hundred and eighty six thousand miles per second, it takes twenty six hours to go and then acknowledgement comes twenty six hours later, and it keeps changing.

Our deepest space network system has three antennas set in Basku in California, Madrid in Spain and Canberra in Australia. So when the earth spins, we have 24/7 coverage. Why do we do this? When I fly for a long ride, if my neighbour sitting next to me somehow learns that I am a physicist working for NASA, ten out of nine times, irrespective of gender, colour of skin, religion or ethnicity, they ask, "Are we alone? Is there life in outer space?" Looks like the question, "Are we alone?" is somehow genetically quoted into us. It's fascinating! It's interesting to note as to why we call this a solar system and not an earth system!

For nearly a few million years we believed the earth was flat, we were at the centre and everything and the universe spun around us. Normal people, also called Homo Sapiens; in Latin, homo is "man" and sapien is "wise" – "wiseman"- were very egocentric and less tolerant, so they thought everything was around us; but some people thought otherwise. Some thought there were other worlds, and they were put to silence very quickly by execution. Aren't we glad we live in more tolerant times today? We can say, "It's flat", "It goes around or not", "I don't believe in it", etc. People may argue with you, but not get physical. I'm going to talk about different types of space-craft we use, one example for each satellite class, such as, Voyager, Cassini, Phoenix and Curiosity. Then I'll talk about the hunt for other earth-like planets and recent developments in the search for life in our solar system and beyond.

JPL was formed by the California Institute of Technology in 1936 as a graduate student experiment with the involvement of six students. JPL gave the first orbiting spacecraft called Explorer 1 to the United States of America in 1958. The first two Russian spacecraft were Sputnik 1 and Sputnik 2. We have about 9000 staff, located in Pasadena in California at the foot of St. Gabriel Mountain. In 1940, JPL's first claim to fame was the development of something called jet assisted takeoff, during the World War II, for planes to takeoff at very sharp angles, so in enemy territory it was very helpful. In 1950, they developed the first guided missile for the United States Army, and in 1958, they designed the first orbiting satellite called the Explorer. Today, we have thirty one robotical spacecraft, two beyond the solar system and the balance twenty nine are around different planets. The four types of spacecrafts we use today are: Flybys, Orbiters, Landers and Rovers. Sometimes planets align; in that case it is much more cost effective to send one satellite to observe few planets. It happens once in one hundred and seventy six years; they are the major plants: Jupiter, Saturn, Uranus and Neptune. Voyager is a Fly-by. We built two voyagers, Voyager 1 and Voyager 2; and launched them to Jupiter in 1976. Jupiter, the largest planet, is a thousand times larger than the earth and does not have a terrestrial land; instead it's just a gas bowl with many moons. Before sending the Voyager, we knew of only four moons, discovered by Gailio called Galilian Moons, bigger than our moon, but Voyager 1 discovered fifty four moons, so Jupiter has fifty eight moons. It's fascinating! In one of the Galilean moons, we observed a big volcanic eruption. This was the first time we observed a volcanic eruption beyond the earth. All the planets of the solar system are on one plain called the Solar Plain.

Voyager 2 was sent two weeks behind, in case something goes wrong with Voyager 1. Saturn is a magnificent planet, also a gas bowl. It has a fantastic ring system, first observed by an Italian astronomer. Its density is so low, if you can take this serene and beautiful Saturn to the ocean, it will float. Its rings are formed with ice particles; some are like sand pebbles and some are big chunks of ice, as big as ten meters. Close to about sixty moons were observed in Saturn; the biggest moon is Titan, at which temperature is very low and ice water is frozen. Another moon of Saturn is called Enceladus. It is a very small moon covered in ice and it has water-rivers. Hence, it has a lot of interest. Voyager is very interesting. Professor Carl Sagan, Professor of Cornell University, encouraged JPL to put a message if there is any intelligent life elsewhere. So we made a copper record quoted in gold with greetings from fifty five languages including "Ayobowan", one hundred and fifteen pictures, a variety of natural sounds of birds, whales, giraffes, lions, etc., and also President Carter's and the then UN Secretary General Waldheim's message, classical and western music, and also we put a needle and sign language if intelligent life captures it they will figure out how to play it. We put a map of the solar system so that they would know from which planet it came from. We also put sign language indicating where we are and where it came from. It is hoped for someone to find it; similar to in early days when people got lost in the ocean or stranded on an island, they would put a message to a bottle, hoping someone would find it.

The first Lander was launched to Mars in 1975. An Italian astronomer found canals in Mars. So, Hollywood movie makers hypothesized Marshians; little men with big heads, complied with Darvin's theory of evolution. We believed Martians were more brainy. Of course now we know Mars does not have intelligent life, but there could be microbian life.

Curiosity is the largest rover we built; it has ten instruments, cameras and very powerful lasers, which would analyze signals coming from vapour to find out what kind of minerals it has. Curiosity has been working on Mars since 2012, and we are building the next one called Mars 2020, and it will be launched in 2020, it will take eight months to go to Mars. We will launch when Mars and the earth are close, so it doesn't have to travel across the solar system, which would otherwise take years. Mars and earth get close every other year. We are about one hundred million miles away. One Martian year is two earth years. We landed on a crater with a five kilo metre high mound. Why did we select this location? From previous Rovers, Landers and Orbiters there is evidence that Mars has running water. We know on earth, life was formed as soon as it had water. So we thought if Mars has running water, this crater could have water. We wanted to explore whether there are rivers. There were pebbles without jaggered edges, instead they were circular, because for millions of years they would have rolled over. One hundred to two hundred years ago Mars had frozen ice like frozen mud, so in Summer times, it melts. The question is if it had water, what would have happened to it? Scientists believe that when the inner core gets colder and becomes solid, it is called a dead planet as nothing moves, and there is no current and no magnetic field. Therefore, due to blasting of high energy solar wind, the water would have vapourized.

Cassini orbiter launched in 1997 on a journey to Saturn, landed in Venus. Until Cassini, we didn't know Enceladus had rivers. We sent Cassini five miles above the surface of Enceladus. It was a very risky maneuver. We found it has geyzers, everything that a primordial soup needs. Now we know of four places that have water: Mars, Europa (icy moon of Jupiter), Titan and Enceladus. We want to investigate all four. We are very much interested in sending a very specific satellite to Titan to explore the possibility of life. We encourage NASA to fund. Actually NASA funded Europa Clipper Mission last year. Europa

ISBN number: 978-955-0301-48-5

is one of the large moons of Jupiter discovered by Galilio. In Europa ice cracks all the time, but we don't know the cause. By 2020 we are going to find out whether there is life.

In 2009 we launched the Kepler telescope to find extra planets orbiting around our neighbouring stars. So far we have found five thousand planets. Out of three thousand five hundred of them we found only two earth-like planets; and in one we think there is water. We shouldn't get discouraged; as the Galaxy has two hundred and fifty billion stars. So far we have discovered about two hundred and fifty billion galaxies. If each star has ten planets, there are so many planets more than all the words uttered by human beings in the last several million years; probably one hundred thousand planets in our solar system that can probably have life.

Now we are building a lot of big telescopes for astronomy. The current largest telescope is a ten metre telescope located in Hawaii. We think with large space telescopes we can hopefully find life elsewhere, within our neighbourhood and in the near future. When looking for life, we look around for earth-like blue planets (blue for water), medium in size, hovering around. Bigger stars or giant red stars burnout fast as their life time is short. Medium size stars like us survive longer for about a million years and their biological process is lengthy, so there is plenty of time to evolve.

The building blocks of life are carbon and hydrogen. There is plenty of these in the universe found by NASA's Spitzer space telescope developed about twenty years ago, still in space. Life must be around carbon and water base. Life can come in surprises. Look at life on earth? Take a bird, lion, jelly fish, human, snake and a giant tree. To explain a snake to a person from another planet – how do you explain? We have this animal as long as a rod, no hands, can swallow his prey five times bigger than his mouth, can move two feet per second, can kill a person like me in a couple of hours by biting and injecting some proteins allergic to us, etc. So when we look for life, it can be in different forms, but water and carbon base.

Our ancestors, probably two and a half million years ago, never thought we would walk on the moon; escape the gravity of earth; become the second kind of species, not the first generation, etc. What's happening? Scientists are making new life! A couple of years ago, French scientists made an artificial rabbit. They took the glowing florescent of a jelly fish and mixed and made a glowing rabbit or a luminous rabbit. Now homosepians are creating life. Many cultures or societies in the East and mostly the West thought only God can create life. Homosepians can play the role of God, too! A question for you!

ISBN number: 978-955-0301-48-5

VOTE OF THANKS

Dr Upali Rajapaksha

The Conference Chairman,
KDU International Research Conference 2018

Honourable Secretary to the Ministry of Defence, Mr. Kapila Waidyaratne, Keynote Speaker, Professor Mohan Munasinghe, Guest Speaker, Dr. Sarath D. Gunapala, Tri-service Commanders, Members of the Board of Management of KDU, Your Excellencies of the Diplomatic Corps, Vice Chancellors of other State Universities, Senior Officers of the Tri-forces and the Police, Our most valued invited guests, Academic and Administrative Staff of KDU Distinguished Ladies and Gentlemen.

It is my privilege to propose the vote of thanks on this occasion. An event of this magnitude cannot happen overnight. The wheels started rolling months ago. It required planning and a bird's eye view for detail. I have been fortunate enough to be backed by a team of motivated and dedicated colleagues, who were willing to take on the completion of tasks beyond their comfort zones.

It is with pride I announce that we received more than 573 manuscripts, from local and international authors, and approximately 350 of them are published. Moving with the times, this year's conference offers great opportunities to presenters, such as the ability to deliver presentations via Skype; and to upload Google Scholar in order to generate individual H-indexing citations.

It is with utmost pleasure I announce that we have also given many opportunities to internal and external undergraduates to share their research findings at our conference, as either poster or oral presentations.

On behalf of KDU, or let me call it fraternity of the one and only Defence University of Sri Lanka, I extend very hearty thanks to our Chief Guest, Honourable Secretary to the Ministry of Defence, Mr. Kapila Waidyaratne; for gracing this occasion. The Support we received from the Ministry of Science & Technology and Bank of Ceylon was immense.

It is my pleasure to acknowledge our gratitude to the Guest Speakers, Professor Mohan Munasinghe and Doctor Sarath Gunapala, for sharing with us their findings and opinions. We are all inspired by your great words. You are an enormous pride to our motherland.

My special thanks go to our Vice Chancellor, Deputy Vice Chancellor Defence and administration and Deputy Vice Chancellor Academic, for your consistent guidance throughout this journey.

Ladies and gentlemen, we thank you for being with us this morning.

Have an inspirational and fruitful day!

Plenary Session

ROLE OF DATA SCIENCE IN DIGITAL TRANSFORMATION OF PUBLIC SECTOR

Terrence Perera
Assistant Dean – Academic Resources,
Sheffield Hallam University, United Kingdom

As digital technologies are rapidly becoming an integral part of our daily lives, every industry is racing to embrace digital transformation as a vehicle to deliver better products and services to their consumers. Public sector organisations are no exception; it is universally accepted that only digital transformation could radically enhance services to public and lower overall operating costs. Within the context of digital transformation, data sciences play a key role by enabling organisations to understand their customers better and provide unmatched customer experience. For example, companies like Amazon, Disney and Uber rely heavily on data analytics to continually improve and innovate their services. This presentation aims to emphasise the role of data sciences in digital transformation projects. Following a brief introduction to key technologies, a wide range of case studies are presented to highlight the use of data science technologies in digital transformation projects. The presentation will also discuss challenges and opportunities of digital transformation projects in the public sector.

Keywords: data science, digital technologies, industry

INTRODUCING STEM EDUCATION IN STATE SCHOOLS ACROSS SRI LANKA

Ashu Marasinghe
The Parliament of Sri Lanka
ashu@parliament.lk

This report presents a proposal for the implementation of STEM (Science, Technology, Engineering and Mathematics) - education across state schools in Sri Lanka. The project partnered with SCCIP Japan Company Limited (SCCIP) and Kaatsu International University(KIU) – STEM Center will manifest the importance of a curriculum based on Robotics Education for students from grade one to grade ten in state schools in order to create a school culture where the significance of STEM education is highly recognized and valued. The project aims to consolidate the skills of students such as critical thinking, application of information gained through experience and reasoning and integration various education disciplines to solve authentic problems thereby, fostering innovation and entrepreneurial spirit among the young work-force of the country. This project will motivate for a strong robotics curriculum in order to encourage STEM education among school students. Robotics education is of a substantial platform to influence students as this will utilize student imagination for the process of designing, building and testing solutions for real-world problems. The novel curriculum will fulfill the needs of economic, scientific and technological developments in the contemporary world of the fourth Industrial Revolution. This will strongly focus on student preparing to enter the job market with the necessary skills set which is an inadequate factor in the current education curricula.

The initial pilot project will commence with the launch of the robotics-based curriculum (developed by SCCIP) to a selected group of state schools across Sri Lanka. The schools include Royal College, Ananda College, Vishaka Vidyalaya, Sangamitha Balika Vidyalaya, Hemali Balika Vidyalaya and Darussalam Maha Vidyalaya. SCCIP intends to provide a STEM Robotics Curriculum for each grade, a STEM Robotics trained teacher and required STEM Robotics Kits and Software. This pilot project would encourage students to participate in Robotics competitions actively, conduct student-teacher exchange programs, summer camps & tours, etc. that will show cast Japanese advanced Technology and to develop the knowledge, skills & habits of mind associated with STEM disciplines by adopting an interdisciplinary & applied approach.

Keywords: School education, Encourge Students, Robotics Curriculum

ENGINEERING HIGH PERFORMING SYSTEMS

Malith Jayasinghe
Director of Platform Architecture, WSO2

The performance of a system is often characterized by its throughput (ability to perform defined sets of tasks within a given period of time), response time (latency) and scalability. Performance is one of the main differentiators that determines the competitiveness of a product in the market. There has been a growing interest in developing high performing systems, particularly in the recent past. Building such high performing systems is a challenging task due to numerous reasons such its the complexity and subjectivity. This paper discusses the core principles and guidelines for developing these high performing systems.

Keywords: performance, principles, complexity

ARTIFICIAL INTELLIGENCE FOR NATIONAL DEVELOPMENT: ADDRESSING REAL WORLD PROBLEMS

HL Premarathne
Former Senior Lecturer, University of Colombo School of Computing Sri Lanka

Artificial Intelligence (AI) is a branch of computer science dealing with the simulation of intelligent behavior in computers. Nearly 75 years of active research on AI, the achievements include the favourable results in several areas such as robotics, code writing, agent technology, ubiquitous computing, IoT etc. However, differences between humans and machines such as emotions, understanding and consciousness require drastic improvements. In addition to commercially motivated products with a flavor of AI, many issues related to the grass root level of a society, especially in the developing world need to be resolved. Close collaborative work among the researchers and the industry will be more productive towards national development.

Keywords: Artificial Intelligence (AI), collaborative research

Technical Sessions

RICE YIELD ESTIMATION USING FREE SATELLITE AND FIELD DATA

Tharshini Shanmugam[1] and TL Dammalage

Department of Remote Sensing and GIS, Faculty of Geomatics,
Sabaragamuwa University of Sri Lanka, P.O. Box 02, Belihuloya, 70140, Sri Lanka
[1]*tharsha0012@gmail.com*

Abstract- An effective pre-harvest rice yield estimation method is truly significant for the assessment of seasonal rice production in terms of strategic planning purposes. In Sri Lanka, a conventional method named crop-cut survey is used to estimate seasonal rice production, yet it fails to forecast rice yield before the harvest as it is conducted during the harvest. Therefore, this study is focused on identifying cultivated paddy lands and forecasting rice yield using free satellite data. Landsat 8 OLI/ TIRS images (30m spatial resolution) from Earth explorer and 8-day composite images (250m spatial resolution) from Moderate Resolution Imaging Spectro-radiometer (MODIS) sensor on board NASA EOS Terra/Aqua satellite were used from 2014 to 2017. Paddy cultivated lands were identified by land cover classification by using field training samples and Landsat 8 OLI/ TIRS data. In addition, the temporal change of Normalized Differenced Vegetation Index (NDVI) for paddy and forest was also analyzed to validate the classification. The observed minimum accuracy of the land cover classification out of the tested four (4) seasons was 99.4%, and the minimum kappa coefficient was 0.9916. The correlation coefficient between reference net harvested paddy area and paddy cultivated area identified by Landsat 8 is 0.93. Linear and exponential yield forecasting models proposed by Sirisena, et al. (2014) for Kurunagala district were validated and tested based on NDVI and EVI2 vegetation indices obtained through MODIS (MOD09Q1v006) surface reflectance image of Polonnaruwa District. The comparison of the estimated yield with national statistical records, both NDVI and EVI2 based models, provide more reliable estimations about 80 days after the transplanting of each season, but, EVI2 based model (derived at 80 days) gives more reliable estimations than NDVI based model with 86.37% of average accuracy. Therefore, seasonal rice yield can be successfully forecasted one month prior to the harvest time using EVI2 based model in the Polonnaruwa district.

Keywords- EVI2, Landsat 8 OLI/TIRS, MODIS, NDVI, Rice yield

I. INTRODUCTION

Paddy occupies approximately 37 % (0.77 million ha) of the cultivated land area of Sri Lanka. It is cultivated during two major seasons; Yala and Maha. The periods of the commencement of each season are uncertain as the commencement of Yala season varies from, end of March to mid of May, and Maha season from end of September to mid of December. The majority cultivates in Maha season. Sri Lanka produced nearly 2.9 million tons of rice in the last season of cultivation (2015/2016). There was an overproduction in that season, and therefore farmers in many parts of the country were angered and disappointed over the failure to set up an adequate mechanism to market their paddy. In the 2016/2017 Maha season the production plummeted as low as 1.7 million tons of rice, and there was food shortage.

Also rice yield estimation in Sri Lanka is based on conventional techniques of data collection for crop and yield estimation based on ground-based field visits and reports. It is time-consuming, subjective, and prone

to significant discrepancies as a result of insufficient ground observations that cause poor rice production assessment. The outcomes are usually made available to the government and public after several months of the harvesting of the rice, and thus not useful for food security purposes, and It is costly, depending on the survey areas. Mostly, the data become available too late for appropriate actions to be taken to avert food shortage. Therefore, we need to estimate rice before the harvesting of the rice.

According to that, satellite remote sensing was widely applied and it was recognized as a powerful and an effective tool for identifying agriculture crops. The possibility to estimate seasonal crop yield before a specific time period of harvesting is vital to take precautions regarding seasonal crop production. Specially, it is of paramount importance in strategic planning and decision making regarding the food security and facilitation of safe harvest storages. On the other hand, proper import (in shortfall case) or export (in surplus case) policies can be made based on such reliable yield estimations (Noureldin et al., 2013). Since there are such considerable advantages related to a pre-harvest yield estimation method, it is very important to focus on modern science and technology in developing reliable yield forecasting models. Satellite Remote Sensing can be successfully applied to this research area as it is powerful and effective in estimating and forecasting crop yields (Ferencz et al. 2004).

Use of satellite data for crop classification and crop yield estimation has a long history. Primarily, for paddy rice classification, existing studies mainly use two types of sensors: Visible and near-infrared (NIR) sensors: such as Landsat, MODIS, and Sentinel-2, Radar sensors: such as ALOS, Radar Sat, and Sentinel-.1. There is very limited study to use satellite data for paddy rice yield estimation at the field level. The major bottlenecks include: Lack of field-level crop yield data (thus current project is very critical) (Lobell et al., 2014). Lack of continuous satellite data. Due to the lack of data, the methodology improvements are slow. Cloudy issues for the visible/NIR sensors (Guan et al., 2012; Gao et al., 2015). Though methodology progress for paddy rice is relatively slow (Nelson et al., 2014), visible and near-infrared (NIR) sensors are most available data. However, it is hard to use it. For classification: a few usable images during the growing season sometime suffice the needs. However, for yield-estimation, we need both: high spatial-resolution for field-level information. High temporal-resolution for growth condition or phenology and we simple donot have the free data at both high resolutions.

The study aims to improve the MODIS images-based rice yield estimation model with the Landsat 8 OLI/TIRS images and determine the best age of paddy plants for yield forecasting in Polonnaruwa, to identify the paddy cultivated area based on Landsat 8 OLI/TIRS images, and to validate the MODIS images based yield estimation model for the study area and Prediction of the rice yield.

II. STUDY AREA AND MATERIALS

A. Study Area

The study area was Polonnaruwa which is located in the North Central Province, Sri Lanka, at 7° 56 latitude 81°0 longitude (Figure 1). Polonnaruwa is the best area for rice production in Sri Lanka and it has approximately 85,505 acres (34,629.525 ha) of rice fields (Irrigation Department of Sri Lanka 2009). Due to the wideness of the area, this farming system can be easily monitored by remote sensing.

Figure 1. Spatial extent and location of the study area in Sri Lanka with a subset of the Landsat 8 true color bands combination, acquired in 2017.

B. Satellite and Rice Yield Data

Free satellite images Landsat 8 OLI/TIRS level 1 collection 1 were downloaded from the USGS Earth Explore Web site from year 2014 to 2017(6 season) images. This product consists of 30m spatial resolution and there is no proper temporal resolution. MODIS Surface Reflectance 8-Day L3 Global 250m (MOD09Q1) Product was used in this study from 2014 to 2017.MOD09Q1 provides Bands 1 and 2 at 250-meter resolution in an 8-day gridded level-3 product in the Sinusoidal projection. Each MOD09Q1 pixel contains the best possible L2G observation during an 8-day period as selected on the basis of high observation coverage, low view angle, the absence of clouds or cloud shadow, and aerosol loading. Science Data Sets provided for this product include reflectance values for Bands 1 and 2, and a quality assurance rating. Rice yield data and other required paddy statistics from 2014 to 2017 were obtained from Census and Statistics Department, Sri Lanka.

In order to achieve the major objective of the study, a method was suggested to identify paddy cultivated lands in each growing season using Landsat 8 OLI/TIRS data and MODIS satellite data. In Landsat 8 OLI/TIRS images Radiometric and FLAASH atmospheric corrections were conducted. Clouds were removed. There were three science data sets in MODIS satellite images which were 250m surface reflectance band 1 (620-670 nm), 250m surface reflectance band 2 (841-876 nm) and 250m reflectance band quality data sets.

III. METHODOLOGY

A. Vegetation Indices

NDVI: The Normalized Difference Vegetation Index (NDVI) gives a measure of the vegetative cover on the land surface over wide areas. Dense vegetation shows up very strongly in the imagery, and areas with little or without vegetation were also clearly identified. NDVI calculated by using NIR band-5(0.845-0.885µm) and Red band-4 (0.630-0.680µm) In Landsat 8 OLI/TIRS images. NDVI temporal change of the paddy and forest was derived. Temporal pattern of paddy cultivated area was identified through that graph.

MODIS images consist of two bands to calculate NDVI. Both bands were used as green vegetation displays strong absorption in the red part of the spectrum (reflectance of around 3-5%) and weak absorption in the NIR part (reflectance around 40 - 60%) (Gitelson, 2004). In MOD09Q1 images calculated NDVI by equation 1 using 250m surface reflectance band 1 (620-670 nm), 250m surface reflectance band 2 (841-876 nm).

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (1)$$

EVI2: The Enhanced Vegetation Index (EVI) (Huete et al., 1999) was developed as a standard satellite vegetation product for the Terra and Aqua Moderate Resolution Imaging Spectro radiometers (MODIS). EVI provides improved sensitivity in high biomass regions while minimizing soil and atmospheric influences. The major limitation of EVI is, it utilizes blue band in addition to the red and near-infrared bands. Jiang et al. (2008) developed and evaluated a 2-band EVI (EVI2), without a blue band, which is synonymous with the 3-band EVI, particularly when atmospheric effects are insignificant and data quality is good. EVI2 index of the MOD09Q1 images calculated.

$$EVI2 = 2.5\frac{NIR - RED}{NIR + 2.4RED + 1} \qquad (2)$$

B. Data Collection

The field route was selected through Google earth application before commencing the field work. The centre of the study area, which consists of more paddy fields was selected for further purposes. The field work was planned for one day. GPS co-ordinations of the paddy fields and the forests were collected through CT Droid Sri Lanka application in mobile phones. 50 paddy samples were collected. Ground truth data was used for the classification. The identification of paddy area and the collection of 20 samples of remaining places were done by using Google Earth application. The areas in terms of percentage and hectares were also computed. Accuracy assessment was captured.

The paddy cultivated area could be identified through the basic statistics of the maximum likelihood classification, and it was compared with the information of census and statistics department (reference data). As mentioned above, there are two main paddy cultivation seasons in the study area, yet, the commencement of the relevant seasons is uncertain. Therefore, several composites (covering the whole season) were taken for each season in terms of time series analysis.

Then NDVI values for all images were calculated using equation 1. Before calculating correlation coefficients, NDVI temporal profiles were smoothed using "moving average" method. Correlation coefficients between NDVI changing pattern of normal paddy cultivation and smoothed profiles were calculated. In this respect, 10 correlation coefficients were calculated for one smoothed

NDVI temporal profile and 10 correlation coefficients were also calculated. Then an appropriate threshold value for correlation coefficient was selected to extract cultivated paddy pixels by equalizing the total paddy area identified through algorithm and census data. The pixels which included correlation coefficient values higher than threshold values, were classified as paddy cultivated lands. A few pixels (identified as paddy) were randomly selected for verification of the classification result. In this process, randomly selected paddy pixels were compared with Google earth and field data to determine whether those pixels are truly paddy or not. This method was applied for the other years of the seasons for identify cultivated paddy lands in the study area. Paddy cultivated area was identified through Landsat 8 and MODIS images. Both extents were compared with reference to data and also check whether the areas are truly cultivated or not with the training samples and Google earth application. Finally get the cultivated paddy extent from analysing those extends.

C. Validation of Model Equation and Yield Prediction.

Dammalage et al. 2017 has proposed several yield estimation models for Kurunagala district using MODIS data. This study investigates the use of the same models to estimate the rice yield in Polonnaruwa district. For validation correlation coefficient from MODIS images was used as the one of the input quantity, and also paddy area from both Landsat 8 images and MODIS images were used to find out which is most suitable for this analysis. According to that, the cultivated paddy area identified through Landsat 8 OLI/TIRS images was widely used for research purposes. After validating and checking the accuracy of each model, the most accurate one will be used for yield prediction in Polonnaruwa district.

IV. RESULTS AND DISCUSSION

A. Identification of Cultivated Paddy Fields

Figure 2 shows the temporal pattern of the cultivated paddy and forests identified through 2017 yala season images.

Generally yala season is from May to August, (not certain). Therefore the images of middle of April - end of August seasons were used. There were paddy fields identified in May. NDVI values of the paddy was increased near NDVI values of forest. In July high NDVI value was occured (after

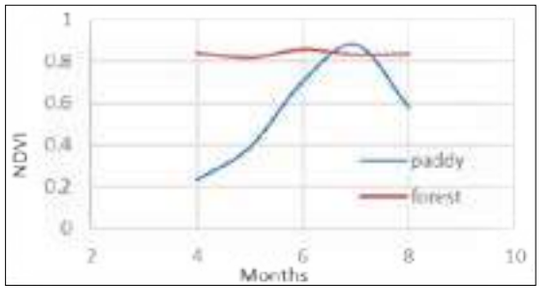


Figure 2.The cultivated paddy lands identification process by temporal change of paddy and forest derived from Landsat 8 images

80 days). However, the utilisation of Landsat 8 images in identifying cultivated paddy fields was unsuccessful due to the low temporal resolution of those images. Then for the land cover classification both (training samples and google earth samples) ground truths were used. Figure 3 shows the classified image over study area.

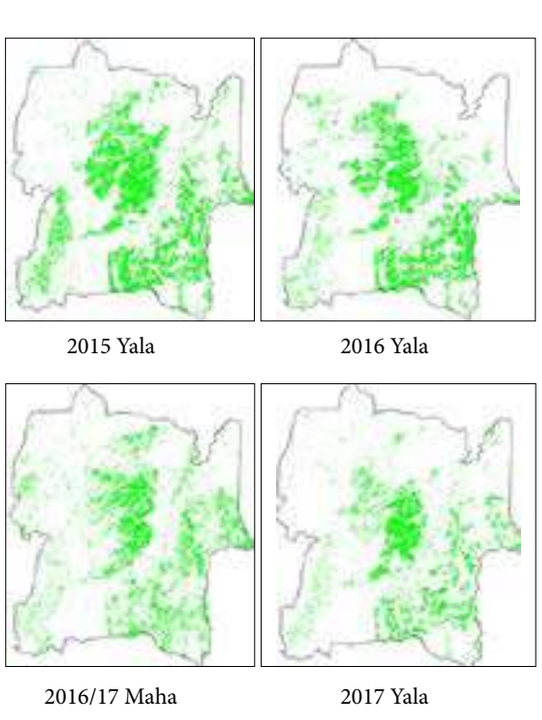


Figure 3. Landsat 8 classified images. (Green colour indicates paddy cultivatedareaFigure 4. The relationship between rice age and Determinant of Coefficient (R2) values of linear yield forecasting models

Only by using the training samples the identified cultivated paddy extent could not be checked. Still have the problem about the identified paddy extend by Landsat 8 image. But according to the time allocation for the analysis it couldn't find out correctly. After that, compare those images with ground truths (training samples). This method was applied for all the other seasons to identify cultivated paddy lands in the study area. Before calculating correlation coefficients, NDVI temporal profiles were smoothed using "moving average" method. Correlation coefficients between NDVI changing pattern of normal paddy cultivation and smoothed profiles were calculated. Then an appropriate threshold value for correlation coefficient was selected to extract cultivated paddy pixels by equalizing the total paddy area identified by this algorithm and census data. The pixels which included higher correlation coefficient values than threshold values, were classified as paddy cultivated lands. Correlation coefficient (r) has taken a higher value at some points when the specific NDVI time series corresponded to a cultivated paddy land. Required seasonal parameters were obtained based on the number of times that the matching pattern was shifted to right occurring the highest correlation coefficient value. In the above situation, maximum value of the correlation coefficient is 0.90 and it is highly correlated with the matching pattern. Therefore that pixel is classified as paddy cultivated pixel. Using this method, all pixels were classified into two classes as paddy cultivated and uncultivated lands.

According to the table 1, there is no unique or common threshold value which can be used to distinguish cultivated paddy lands from other Land uses. However,

Table 1. demonstrates, the interdependency between the amount of area identified as paddy and the threshold value for correlation coefficient.

Season	Threshold for correlation coefficient	Total amount of area identified as paddy (hectares)	Net harvested area (hectares)
2014/15M	0.8116	65,756	65,891
2015 Y	0.8454	54,513	54,668
2015/16M	0.8990	57,650	57,608
2016 Y	0.8920	49,731	49,754
2016/17M	0.9057	42,306	42,268
2017 Y	0.8615	35,625	35, 585

by changing the shape of the matching NDVI pattern and smoothing and filtering methods of time series data, a unique threshold value can be discovered and it provides food for thought for further studies.

B. Relationship between Vegetation Indices and Rice Yield

A number of rice yield forecasting models had been built up based on NDVI and EVI2 vegetation indices at different age of paddy starting from a rice age of 32 days.

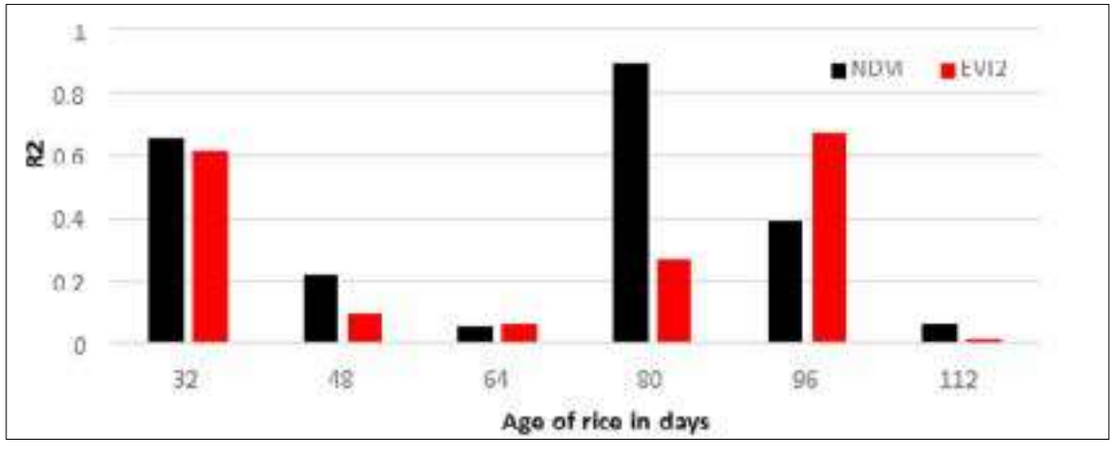


Figure 4. The relationship between rice age and Determinant of Coefficient (R2) values of linear yield forecasting models

Figure 4 shows how determinant of coefficient (R2) values of derived linear models change with the age of the paddy plant. The trend of R2 slightly decreased from the age of 32 days to 64 days, then it has reached its maximum value at the age of 80 days and again it decreased until the harvest time. Based on this statistical analysis, the paddy plant at about 80 days after transplanting has given the best relationship between vegetation indices (NDVI and EVI2) and rice yield. However, in most stages, yield forecasting models which are derived based on EVI2 index have higher R2 values than NDVI based models.

To determine the accuracy and reliability of estimations provided by yield forecasting models, an accuracy assessment has been done for Polannaruwa district over six seasons by using rice yield data provided by the Department of Census and Statistics in Sri Lanka. The accuracy of the estimations is given in table 2 and those results have shown the reliability of yield estimations provided by each model which are based on remote sensing data. Table 2 illustrates the validation of each model over the Polonnaruwa district. There are 5 linear and 5 exponential models. Each and every model consists different accuracy level. According to this validation the correlation between reference data and identified data by using these models were given, NDVI after 80 days gives 0.89 and EVI2 gives 0.27 . Thus, NDVI after 80 days gives an overall accuracy of 77.54% and EVI2 after 80 days model gives an overall accuracy of 83.68%. Therefore, the EVI2 after 80 days model is suitable for rice yield estimation over the study area. Then correlation between the reference yield data and estimated yield data was calculated by using the graphs. X axis determined reference yield data from statistics department and y axis determined estimated rice yield by models. There was good correlation in NDVI after 80 days model, but it has less accurate than EVI2 after 80 days model. Therefore, EVI2 after 80 days based model can used to estimate the rice yield.

B. Languageimprovement needed

Table 3 shows that the comparison of cultivated paddy area. According to that,the area identified by Landsat 8 was in 2

Table 2. Different type of rice yield forecasting models and a comparison of their forecasting results with national statistical data.

Forecaster	Model	Accuracy (model/crop cutting) ×100%					
		14/15M	15 Y	15/16M	16 Y	16/17M	17Y
NDVI after 80 days	$y = 2542x + 2087$	73.2	77.7	78.9	79.3	78.1	80.3
	$y = 2391e^{0.6862x}$	73.2	77.8	79.0	79.4	78.2	80.3
EVI2 after 80 days	$y = 8402x + 174.3$	72.9	88.5	81.2	91.8	78.7	96.5
	$y = 1410e^{2.295x}$	72.9	90.6	81.6	94.4	78.9	99.8

Table 3. Paddy cultivated area comparison

Season	Identified paddy area by Landsat 8 (Hectares)	Identified paddy area by MODIS (Hectares)	Net extent (Reference data) (Hectares)
14/15 M	No image	65,756	65,891
15 Y	67,223.61	54,513	54,668
15/16 M	No image	57,650	57,608
16 Y	54,847.53	49,731	49,754
16/17 M	47,185.47	42,268	42,268
17 Y	36,739.80	35,625	35,585

Table 4: Accuracy analysis of total rice yield production by models

Season	Reference total yield (MT)	NDVI after 80 days model (MT)	EVI2 after 80 days model (MT)	Accuracy (model/crop cutting) ×100%	
				NDVI After 80 days	EVI2 after 80 days
15 Y	280,476	268,470.05	312,556.34	95.70	Over production
16 Y	251,131	219,826.87	261,353.47	87.53	Over production
16/17 M	214,722	187,489.52	189,029.05	87.32	88.03
17 Y	173,595	143,935.52	178,805.26	82.90	Over production

decimal places . And according to the accuracy analysis and other analysis the identified cultivated area by Landsat 8 was more accurate than the other data. But still we have the issue on this data because for the above mentioned reasons.

Table 4 shows the accuracy analysis of estimated total rice yield by different models. NDVI after 80 days model and EVI2 after 80 days model is considered to estimate total production. Therefore, the total paddy cultivated extent taken from Landsat 8 images and estimate total yield. So according to that, table using NDVI model will give an accuracy closer to that of crop cutting survey method. But using EVI2 model gives a higher accuracy than the crop cutting survey method. Accordingly, those two models can be used for estimation if the reference yield data is assumed to be correct, and if need to the accuracy as crop cutting survey can replace this NDVI after 80 days model because one month before harvesting can estimate rice yield near to crop cutting survey accuracy. Therefore, no need to wait until crop cutting survey. Another thing is that the EVI2 based model consists total estimation more than crop cutting survey output. Therefore, it also can use, but have to analyse about whether the total yield production is correct or not. Couldn't say that the reference data was correct, anyhow have to analyse and then can use that model also for further analysis. According to time allocation it couldn't find out.

V. CONCLUSION AND RECOMMENDATION

The algorithm for identifying cultivated paddy fields was developed based on high spatial resolution images of paddy cultivation. The developed algorithm retains the capacity to find the commencing period of paddy cultivation though it alters from one location to another. The impossibility of using Landsat 8 images successfully was a major limitation of the study. Therefore, as this limitation highly influenced the utilization of Landsat 8 images in the identification of cultivated paddy fields, the relevant areas should be visited and observed to collect paddy statistics and field verifications. Nevertheless, the accuracy of this algorithm can be improved by using high temporal resolution images.

Finally, the researchers arrive at a conclusion that the model based on EVI2 values is the best model for rice yield forecasting using satellite imagery (This study was only based on NDVI and EVI2 indices). Therefore, rice yield can be accurately forecasted approximately for one month (4-month paddy) before harvesting.

REFERENCES

Cheng, Q., & Wu, X. (2011). Mapping paddy rice yield in Zhejiang Province using MODIS spectral index. Turkish Journal of Agriculture and Forestry, 35(6), 579-589.

Dammalage T. L., Sirisena P.M.T.S. and Junichi Susaki (2017). MODIS Satellite Data Based Rice Yield Forecasting Model for Sri Lanka: A Pilot Study in Kurunegala District. Asian Journal of Geoinformatics, Volume 17, Issue 3, 2017, pp. 24-33. ISSN: 1513-6728

Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (2013). Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. PloS one, 8(8), e70816.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. Remote sensing of Environment,112 (10), 3833-3845.

SOFTWARE FOR COMPLEX PROCESS AUTOMATION AND STAKEHOLDER RELATIONSHIP: STATE-OF-ART IN HYDROGIS TOOL FOR URBAN FLOOD MANAGEMENT

RMM Pradeep[1] and NTS Wijesekera[2]

[1]General Sir John Kotelawala Defence University, Sri Lanka
[2]University of Moratuwa , Sri Lanka
[1]*rayantha@gmail.com*

Abstract- To make sustainable decisions in policy making/ public decision making, the assisting software should provide sustainable options. This is very important in urban flood management, due to the involvement of number of stakeholder groups. In order to develop such software, a software development effort needs to realise the basic requirements of sustainable decision making, which need more research. The present work attempts to study the available researches for sustainable decision making process in urban flood management and analyse according to the software development profession. The present work utilises a HydroGIS tool development effort, which was developed for urban flood management, to review the literature findings. The study found the importance of understanding the complex-process integration with recipient stakeholders for the development of a sustainable decision making software.

Keywords- Sustainable software, Recipient Stakeholders, HydroGIS tool, Urban Flood Management

I. INTRODUCTION

A. Project Stakeholder Management

Stakeholder Management is the latest knowledge area (KA) of the Software Project Management. The requirements of the stakeholders may change with the project development and such requirement change may lead to make drastic decision of shutting down the project. Therefore, a close monitoring of the project stakeholders is a sine-quo-non in today's project management. However, once the required software is produced to the users/clients, then it considers enclosure (PMBOK® Guide – Sixth Edition, 2017; SEBoK contributors, 2015).

B. Software Sustainability and Recipient Stakeholders

Nevertheless, if the produced software does not assist stakeholders to arrive at a sustainable decision, the software becomes a useless tool to the stakeholders even if it provides technically accurate and feasible answers. Specially, if such decision is made by the governing bodies/ policy makers and negatively affecting the general public, the decision makers tend to make a fresh decision violating the technical guidelines to favour of general public for reduce the resistance. Then these recipient stakeholders who are not direct users of the software but effected through the result of the system, directly influence the sustainability of the software. It has realised that the software developers should identify and provide the facility to users of the system to incorporate recipient stakeholders' requirements whilst the decision making process. At present practise the software developers gather recipient stakeholders' requirements through the end-users who are not capable to clearly express their own requirements too (Becker et al., 2015; Mysiak, Giupponi, & Rosato, 2005; Penzenstadler, Femmer, & Richardson, 2013; Venters et al., 2018).

C. The Multiple Complex Processes Automation

In other view, the decision making process in government/ national scale may contain different complex processes. These complex processes lay on different expert areas which may far away from the software developers' capabilities. Then the code development and testing of such software become more difficult task. Further, when such individual processes required to be changed to satisfy the recipient stakeholders, the developers need to have a clear understanding about process as well as communication between processes and stakeholders.

D. Urban Flood Management

The decision making process in the urban flood management is a one of the prominent scenarios which urges better recipient stakeholder management facility. To arrive to an optimum flood management decision, the governing authorities utilised experts from different areas such as hydrology, town planning and finance. Then they can develop hydrologically accurate, economical plan which match with the existing town plan. But when install the plan on the ground, citizen start to resisting due to interruption to their social, economical and financial practises. Then the governing bodies have to either stop the project or progress with the project align to the citizens, but violating town planning, economical or hydrological decisions. Therefore, the decision support system needs to provide the facility to optimise the solution with incorporating the citizens' requirements at decision making level (Gray, Paolisso, Jordan, & Gray, 2017; Voinov et al., 2016; Weiler & Beven, 2015).

Therefore, when develop such software, developers need to identify (1) recipient stakeholders, their requirements against the scientific decisions, tread offs of both recipient stakeholders and decisions makers and (2) Interaction between different complex processes. Nevertheless, no study has found which analysing both the complex processes handling and recipient stakeholder management on the data and processes sharing perspective.

E. Aim

Then, the aim of the present work is to identify and review the state-of-art in data and process relations between users, recipient stakeholders and different complex processes using a case study of urban flood managing HydroGIS.

II. LITRETURE REVIEW

A. Case Study: Urban Flood Management HydroGIS Tool

As the case study, it selected the stakeholders and processes described in the work of Pradeep and Wijesekera (2011, 2012). Accordingly, it has developed a software which assist local government technical officers (TO) to grant the permission to citizens to carryout land modifications, considering the contribution to the urban flood due modifications. If the required modification is effect on the flood, the TO is allowed to reach a solution with both/ either readjust the modifications and/or incorporate a detention tank to minimise the contribution. For this work, the software developers had to share the data with GIS and Hydrology processes and allow end user (TO at local authority) to optimise the solution by negotiating with recipient stakeholders, the citizens.

B. User roles whilst Integrating Complex Processes in HydroGIS Tool

The main processes involved in the HydroGIS tool are hydrology process and GIS processes. Due to the requirement of automating a hydrological calculation sequence using GIS capabilities, developer has to identify how to integrate GIS and hydrological processes.

At the early stages in 1990s the hydrological calculations and GIS integration carried out using two approaches namely (1) loosely coupling and (2) tightly coupling (Figure 1). In loose coupling approach, hydrological calculations process gets the required parameter values from GIS software processes, manually. Then when it required displaying the results on maps, it has to reproduce the data to GIS. In tightly coupling approach, hydrological process and GIS software are sharing the information required by both hydrology and GIS processes, through software codes.

When formulating these approaches the user involvement made an influence to develop two approaches. The loosely coupling approach is a researcher-oriented which needs more engineering knowledge in bolting hydrology and GIS. The tightly coupling approach needs to facilitate less technical users to perform hydro calculations using GIS environment.(Stuart & Stocks, 1993)

With the development of GIS technology over the time, the attention had being paid to use the GIS capabilities in data analysing and accurate data representation in environmental modelling. Sui and Maggio (1999) describe the integration approaches in four different ways as shown in the Figure 2. The added new approaches were integrating the Hydrology/GIS calculation steps into Hydrology/GIS software tool. It can observe that behaviour and responsibilities of the users in integration of hydrology modelling and GIS capabilities become a

reason to develop four different approaches. However the user profiles for each approach is doubtful as shown in the Table 1. When analysing, it can see that the user is a modeller as well as a software developer.

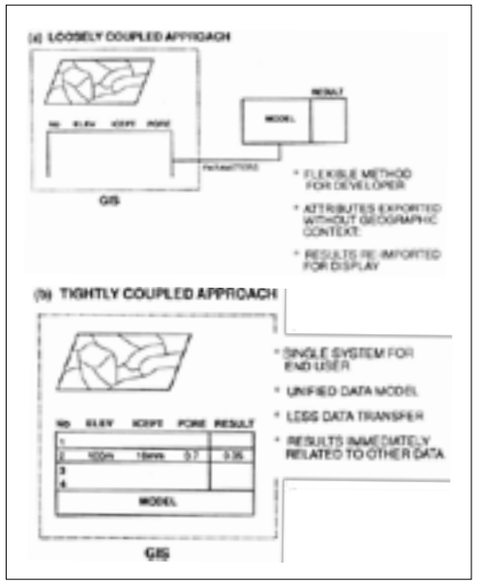


Figure 1. Two Alternative ways of linking a model to a GIS
Source : (Stuart & Stocks, 1993)

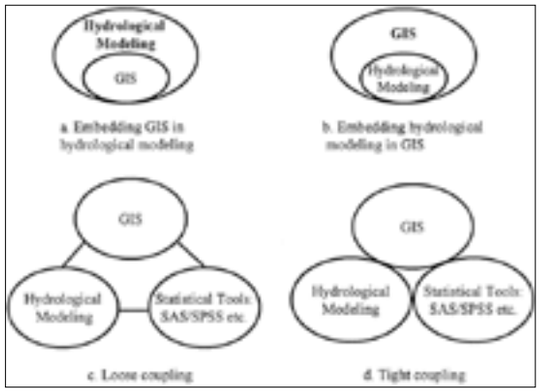


Figure 2. Integrating GIS with hydrological modelling
Source: (Sui & Maggio, 1999)

Further to the user roles in integration, Huang and Jiang (2002) have considered integrating data and/or functions in GIS and hydrology models and summarised the four approaches to three approaches; loose coupling, tight coupling and full coupling as shown in the Figure 3. However the full coupling may either development of software codes to hydrology model processes within the GIS software or development of hydro modelling software with GIS capabilities (Alcaraz, Vázquez-Suñé, Velasco, & Criollo, 2017). Then this full coupling can be considered as "Embedding".

Table 1. User role in Hydro GIS integration approaches – Author Review

Approach	User profile	Disadvantage
Embedding GIS functionalities into hydrological modelling software.	Users are Hydrological modellers, who need GIS as a mapping tool. The requirement of programming skill upgrade the hydro modeller to SW developer/hydro modeller	As hydrological modelling software do not have GIS functionalities, then an intensive programming effort need.
Embedding hydrological modelling in to GIS software.	Unclear whether the hydro modeller or GIS modeller but can be described as GIS software users. Users use inbuilt hydro models in GIS software	GIS functionalities are satisfied. But the hydro model validation is doubtful
Loose coupling	Hydro and GIS software connection carried out by data exchange with less programming. Hence most GIS users and Hydro modellers can use this approach	Data conversion becomes a responsibility of users.
Tight coupling	Users use scripting or general programming language within the GIS to automate the hydro model. Hence user has to be a highly technical person.	Users are allowed to customize user routing, but need to consider the spatial data structure.

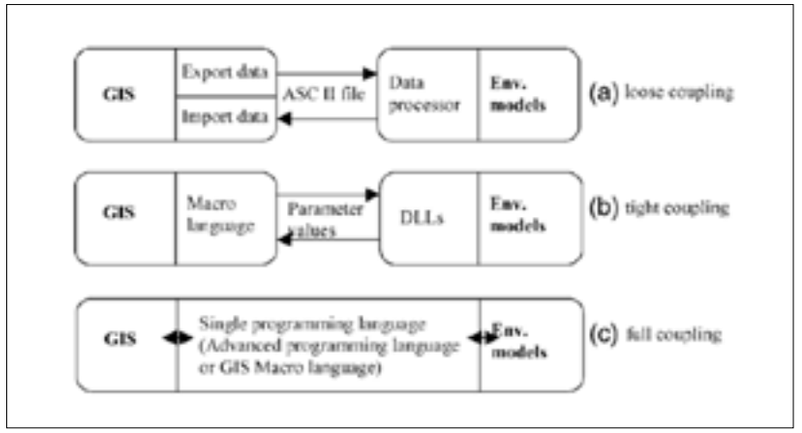


Figure 3. Different approaches to the coupling of environmental models with GIS
Source: (Huang & Jiang, 2002)

C. Stakeholder factor of HydroGIS tool

When consider the different user roles involved in the water resource management decision making in hydro-GIS integrated tool development, an ambiguity arises. To clarify the users, it considered the influential participants in water resource management decision making. Through the outline study, it found a history long discussion about the public participation in decision making which does not considered in Hydro-GIS integration. However public participants are a key interest group of users. The Arnstein (1969) discussion on the different level of engagement of participation of public in decision making which varying from manipulation (non-participation) to citizen-control (fully managerial power), added new knowledge to all decision making disciplines.

In the field of water management, public participation is considered as a key principle. Pioneering Dublin Statement (1992), Water Frameworks Directive (EC, 2000) and the Hague Declaration (2000) recognized the requirement of involvement of stakeholders in each level of water decision making. However application of this principle remained problematic due to decision makers (government) unwilling to participate public, limited/ absent response from the public, low quality response from public and difficulty of conclude the decision making with consistency due to expenditure issues, information hiding from each other or lack of time (Mostert, 2003). As these difficulties public participation become a real challenge that need to manage carefully to arrive to a sustainable water management solution.

In incorporating the public to the water management processes, Henriksen et al., (2009) attempted to involve stakeholders to water resource modelling. They have identified 3 stakeholder groups based on the influence on decision making such as (1) Consultation (opportunity to comment /views), (2) Interaction (allow to advice but decision makers have power to accept or reject) and (3) Engagement (negotiate and engage in trade-offs with traditional power holders). Further they involve the users in determination of the requirement at model study plan and review steps of all the modules such as data and conceptualization, model setup, calibration & validation, simulation and evaluation. However the user role is around "Interaction". This study shows the academic maturity of stakeholder study in water resource management.

However, researches use these stakeholders involvement in hydrology modelling when the watersheds are spread over different nations and cultures. Comair et al. (2014) work is one of such example which stakeholder engagement in water resource management in global context exceeding the trans-boundaries. Nevertheless, the HydroGIS integration is not considered when integrating stakeholders in decision making process.

Table 2. Summary of the User role in HydroGIS integration approaches

Approach	Process integration by[1]	Data Integration through[2]	Author's Review	
			Knowledge required	User & Role
Loose coupling	Users	Files sharing	Spatial Data formats, inputs preparation and output interpretation	Modeller and decision maker: *Use hydro/GIS software for decision making. Data preparation and sharing between processes are done by themselves*
Tight coupling	Users and software codes	Inter software Parameter passing	Software coding knowledge, understand the architecture of both software	Software Developer: *Integrate and develop a system with data preparation and sharing facility between different processes* Modeller/ decision maker: *use the developed system*
Embedding GIS in hydro model	Users using Hydrological Software	Parameter passing within the modules in the software	In-depth knowledge in GIS function automation	
Embedding hydro model in GIS	Users using GIS Software		In-depth knowledge in Hydrology model automation	

[1]Integration by Stuart and Stocks(1993) and Sui and Maggio (1999)
[2]Integration by Huang and Jiang(2002)

III. RESULT AND ANALYSIS

A. Results for HydroGIS Tool User roles whilst Integrating Complex Processes in HydroGIS Tool

Through literatures it can identify four classifications for integrating which can summarize into the three classifications. However integration is setting up for sequential process of GIS and hydrological functionalities which is verified by the modellers. But the user role in this process is unclear. As well all the integration attempts were tried to share the model steps complexities with the different candidate software, such as with GIS software, statistical packages and hydrological software. Author reviewed that the integration is conceptually discussing how this software interacts to perform to get an output from hydrological model. This can be described under two different concepts.

Process integration concept: When carrying out a HydroGIS calculation, the integration can be divided in to four approaches based on the software and user involvement in carrying out processes (Figure 2).

Data integration concept: Aforesaid process can be automated (use software coding to handle the sequence) or can be handle manually (processes carryout using different software by users). Then the data handling responsibility has to be solved and it became an important consideration in integrating.

The data integration classification is based on the data sharing with the hydrology model and GIS software whilst the calculation process. Three approaches of data sharing can be observed when review the literatures, (1) User generates required data using either GIS or hydrological modelling software and share with the counterpart software (2) User operates either GIS or hydrological software, the data sharing is done through intercommunication between software itself (3) User handles the processes steps in a single software, then software perform all the functionalities and pass the data between hydrology and GIS through the developed codes itself. Therefore the two approaches of embedding hydrology model and GIS which shown in Figure 2 (a and b parts), are considered as a single approach as Full coupling in Figure 3.

User knowledge and Role: Foresaid different integrations are based on the approaches made to perform a hydrological model to arrive to a water management decision. Within these processes, users have to use GIS and Hydrology modelling tools. But the level of knowledge required to handle is varying from software development knowledge to tool operating knowledge. Therefore a doubt arises when clarifying the term "user" in the integration. Then a set of users and their roles were reviewed and formulated based on the knowledge requirement in integration. Accordingly, Table 2.0 describes the user roles based on their knowledge and engagement in decision making process.

Then, when study the created user and their roles, it's clear that the integration attempts were made without considering the model development and decision making processes. If the decision making team has all the knowledge such as Hydrology modelling, software development, data management and GIS software handling then integration of hydro-GIS can follow any approach. Nevertheless, always decision making teams consist governing authorities and modellers. Then if the "modeller" carryout all these integration and provides the information required, the decision maker has only to reach a sustainable water management decisions. But to reach the sustainable decision it required the stakeholders' ideas to be considered from model development to decision making. However literature proved that the GIS is a better option to fulfil the communication requirement of this kind of relation. (Jessel and Jacobs, 2005). So incorporating recipient stakeholder is an advanced study than HydroGIS integration approaches. It has to study how different users and their roles interacting with GIS and Hydrology when model development for water resource management.

B. Results for Stakeholder factor of HydroGIS tool

When consider the both the Hydro-GIS integration and stakeholder-hydro modelling relation, the user role of the stakeholders became a problem. The present work review that, if the user role in the process of water management can be solved, then it facilitates to select the most suited hydro-GIS integration approach. Then when solve the user roles it needs to define a clear demarcation of the responsibilities in integration. The responsibility of hydro-GIS integration activities can be only defined among the different stakeholders if it clearly identifies the users' roles in each and every integration point of water resource decision making process.

Therefore in Hydro and GIS integration, the users who are working with the model are called "Modellers". In stakeholder and Hydro Modelling relation the users are the general public and decision makers. Then basically it can identify three different users in the HydroGIS assisted water resource management, such as (1) Modellers (2) Decision Makers and (3) General Public / Recipient stakeholders.

Then integration of processes and stakeholders according to the present literatures is shown in the Figure 4. Description of the processes and data in the figure is shown in the Table 3.

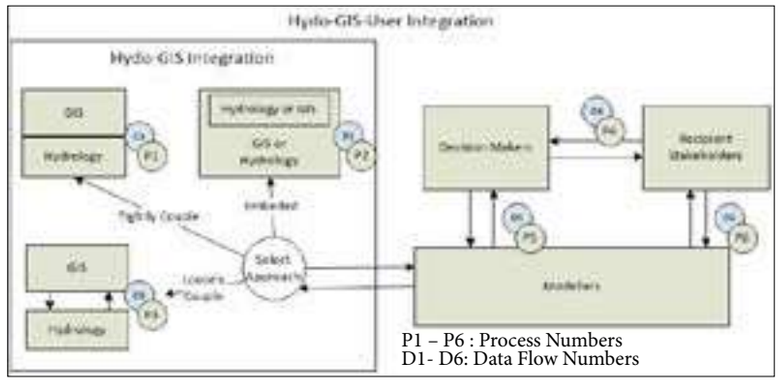


Figure 3. Hydro-GIS-User Integration

Table 3. Involvement of Users in different integration approaches

Process / Data	Description	Modeller	Recipient Stakeholder	Decision Maker
P1/D1	Tight coupling of processes and data	integrate the process and data via coding	No description	Provides rules and regulation to modeller
P2/D2	Embedded processes and data	Embedded one system to other using coding		
P3/D3	Loosely couple processes and data	manually integrate the process and data		
P4/D4	Coupling / embedding decision makers and stakeholders	N/A	1. Either one will adhere to other's requirement / rules 2. Requirements and rules are matching each others 3. Both can negotiate to a conclusion	
P5/D5	Coupling / embedding decision makers and modellers	Integrate hydro and GIS based on decision maker's rules	N/A	Provides rules and regulation
P6/D6	Coupling / embedding modellers and stakeholders	1. Either one will adhere to other's requirement / rules 2. Requirements and rules are matching each others 3. Both can negotiate to a conclusion		N/A

Note: Data and Process integration is shown from D1 to D3 and P1 to P3. D4 to D6 and P4 to P6 show the data and process integration between different users.

Hydo-GIS-User Integration

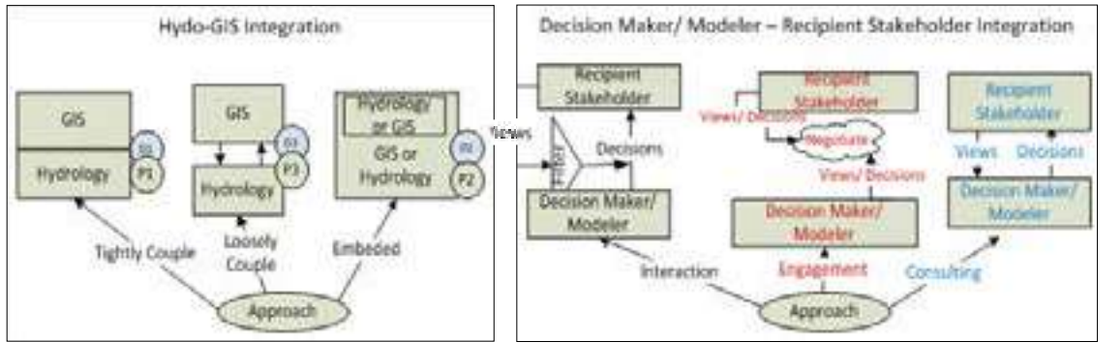


Figure 4. Final Hydro-GIS-Stakeholder integration scheme

Table 4. Stakeholders of HydroGIS tool

Stakeholder	Description	Example
Recipient Stakeholders	The general public who get the benefits/suffers from the decision made	Land owners, citizens
Regulating Stakeholders	The person/s who take decision in development	Local Authority such as Urban council
Institutional Stakeholders	The individuals who technically develop and process the decision making process	Hydro and GIS modellers, tool operators & developers

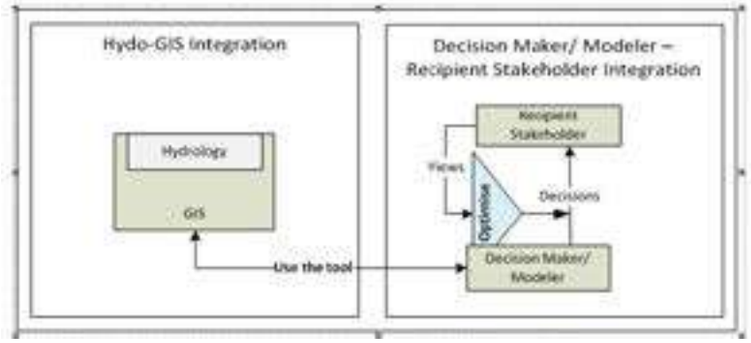


Figure 5. Hydro-GIS-Stakeholder integration schema for Automation

C. Hydro-GIS-User Integration

According to the results it has realised that, recipient stakeholder and decision maker integration has different approaches. Hence in Hydro-GIS- Stakeholder integration, there are two different approaches are to be selected. Then after incorporating the different levels of involvement in decision making, the final picture of the Hydro-GIS-Stakeholder integration is as shown in Figure 4. In this view, the hydrology/GIS modellers are disappeared, but it can identify, for selection of approaches and assisting in stakeholder integration, there should be another group of stakeholders.

Then this additional stakeholder is a group of people which consists of modellers, tool operators and if there is automation, software developers. Therefore the final stakeholders can be grouped and identified as shown in the Table 4.

D. Automating Hydro-GIS-User Integration

Finally the present works analysis the integration requirement of automating the entire processes which can be run by non-technical decision makers/tool operators. Whilst this automation, author realised that the optimization of the recipient stakeholders' requirement can be achieved through "Interaction" option of stakeholder relation. As well considering the well-developed GIS software industry, the present work read the situation to select embedding the Hydrology model in to GIS software as the best option. Then the most suited hydrology-GIS-Stakeholder integration is shown in the Figure 5.

V. DISCUSSION

The present work considers the suited integration scheme for hydro-GIS-stakeholder should be similar to the schema shown in in Figure 5, when automating a hydro-GIS process which is to be utilised in urban flood management decision making with recipient stakeholder's requirements optimization.

The integration consists of two different areas which are; (1) Hydro model-GIS software integration; and based on the maturity of GIS software industry, the work proposed to integrate hydro model in to GIS software and (2) Recipient stakeholder – decision maker integration. As well considering the importance of sustainable solution provide for urban flood management, work proposed to select "Interaction" option of stakeholder-decision maker integration.

Then when automating, software developer can utilise the complex GIS processes in GIS software for carryout the processes of hydrological calculations, trusting the accuracy of the base GIS software outputs.

As well to facilitate the optimization of recipient stakeholder, software developer has to be more emphasises on customising a trial-and-error facility for inputs and outputs to and from the models.

VI. CONCLUSION

The software sustainability is depending on the way the software assists decision makers/policy makers to arrive sustainable decision in urban flood management.

Then to provide such facility, the software should capable to facilitate recipient stakeholder's requirements optimization with decision maker's requirements.

Hence the software developer need a better understanding about the processes & data integration and recipient stakeholder influence on sustainable decision making.

Therefore when develop software for policy making or public decision making which uses multiple complex processes, the development effort should realised the process integration limitations and recipient stakeholder influencing inputs and outputs in the planning stage of the software development life cycle.

The results are formulated through evaluating the experience in HydroGIS tool development activities against the literature review. Then the finding is valid and limited to hydro-GIS tool development for urban flood management. However based on the literatures reviewed the upper limitation can be increased to multi-stakeholder water resource management decision making.

The present work highlights the importance of studying the "User" as not only software operator but also recipient stakeholders in the term of sustainability of the software use in practical scenario.

ACKNOWLEDGMENT

Author pays gratitude to Ms. Malani Herath of University of Moratuwa for assisting explore recipient stakeholder concept.

REFERENCES

Alcaraz, M., Vázquez-Suñé, E., Velasco, V., & Criollo, R. (2017). A loosely coupled GIS and hydrogeological modeling framework. *Environmental Earth Sciences, Springer,* 76(382). https://doi.org/10.1007/s12665-017-6709-3

Arnstein, S. R. (1969). A Ladder Of Citizen Participation. *Journal of American Institute of Planners,* 35(4), 216–224. https://doi.org/10.1080/01944366908977225

Becker, C., Chitchyan, R., Duboc, L., Easterbrook, S., Penzenstadler, B., Seyff, N., & Venters, C. C. (2015). Sustainability Design and Software: The Karlskrona Manifesto. *Proceedings of the 37th International Conference on Software Engineering, 2,* 467–476. https://doi.org/10.1109/ICSE.2015.179

Comair, G. F., McKinney, D. C., Maidment, D. R., Espinoza, G., Sangiredy, H., Fayad, A., & Salas, F. R. (2014). Hydrology of the Jordan River Basin: A GIS-Based System to Better Guide Water Resources Management and Decision Making. *Water Resources Management,* 28(4), 933–946. https://doi.org/10.1007/s11269-014-0525-2

EC. (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Parliament,* L327(September 1996), 1–82. https://doi.org/10.1039/ap9842100196

Gray, S., Paolisso, M., Jordan, R., & Gray, S. (Eds.). (2017). *Environmental Modeling with Stakeholders.* Springer. https://doi.org/10.1007/978-3-319-25053-3

Henriksen, H. J., Refsgaard, J. C., Hojberg, A. L., Ferrand, N., Gijsbers, P., & Scholten, H. (2009). Harmonised principles for public participation in quality assurance of integrated water resources modelling. *Water Resources Management,* 23(12), 2539–2554. https://doi.org/10.1007/s11269-008-9395-9

Huang, B., & Jiang, B. (2002). AVTOP: A full integration of TOPMODEL into GIS. *Environmental Modelling and Software,* 17(3), 261–268. https://doi.org/10.1016/S1364-8152(01)00073-1

Mostert, E. (2003). The challenge of public participation. *Water Policy,* 5, 179–197.

Mysiak, J., Giupponi, C., & Rosato, P. (2005). Towards the development of a decision support system for water resource management. *Environmental Modelling and Software,* 20(2), 203–214. https://doi.org/10.1016/j.envsoft.2003.12.019

Penzenstadler, B., Femmer, H., & Richardson, D. (2013). Who is the advocate? Stakeholders for sustainability. *In roceedings of the 2nd International Workshop on Green and Sustainable Software (GREENS '13)* (pp. 70–77). San Francisco, California: IEEE Press Piscataway, NJ. https://doi.org/10.1109/GREENS.2013.6606424

PMBOK® Guide – Sixth Edition. (2017). *PMBOK® Guide – Sixth Edition. Stakeholder Needs and Requirements.* Retrieved from https://www.pmi.org/pmbok-guide-standards/foundational/pmbok/sixth-edition

Pradeep, R. M. M., & Wijesekera, N. T. S. (2011). *Tool Development for Dynamic GIS Application for Watershed Management Case Study of the Thimbirigasyaya Ward of the Colombo Municipal Council Area Tool Development for Dynamic GIS Application for Watershed Management.* University of Moratuwa.

Pradeep, R. M. M., & Wijesekera, N. T. S. (2012). GIS Tool to Mitigate Urban Flash Floods Through Capturing Land Allotment Modifications and Incorporating Conceptual Detention Pit. In *ITRU 2012.* Moratuwa: Universiry of Moratuwa.

SEBoK contributors. (2015). Stakeholder Needs and Requirements. Retrieved July 22, 2018, from http://www.sebokwiki.org/w/index.php?title=Stakeholder_Needs_and_Requirements&oldid=51430

Stuart, N., & Stocks, C. (1993). Hydrological modelling within GIS: an integrated appraoch. In *Application of GIS in Hydrology and Water Resources: HydroGIS 93* (pp. 319–329). Vienna: IAHS.

Sui, D. Z., & Maggio, R. C. (1999). Integrating GIS with hydrological modeling: Practices, problems, and prospects. *Computers, Environment and Urban Systems,* 23(1), 33–51. https://doi.org/10.1016/S0198-9715(98)00052-0

Venters, C. C., Capilla, R., Betz, S., Penzenstadler, B., Crick, T., Crouch, S., … Carrillo, C. (2018). Software sustainability: Research and practice from a software architecture viewpoint. *Journal of Systems and Software,* 138, 174–188. https://doi.org/10.1016/j.jss.2017.12.026

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., … Ramu, P. (2016). Modelling with stakeholders - Next generation. *Environmental Modelling and Software,* 77(January), 196–220. https://doi.org/10.1016/j.envsoft.2015.11.016

Weiler, M., & Beven, K. (2015). Do we need a Community Hydrological Model? *Water Resources Research COMMENTARY,* 51, 7777–7784. https://doi.org/10.1002/2014WR016731.Received

NEW CUSTOMER CHURN PREDICTION MODEL FOR MOBILE TELECOMMUNICATION INDUSTRY

LLG Chathuranga[1], RMKT Rathnayaka[2], and HI Arumawadu[3]

[1]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

[2]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

[3]Dialog Axiata PLC, Colombo, Sri Lanka

[1]*chathurangagihan39@gmail.com*

Abstract- The present Sri Lankan telecommunication industry remains extremely dynamic by constantly changing the landscape of new services, technologies, and carriers. Thus customers have more choices. So, predicting customer churn is one of the most challenging targets in the telecommunication industry today. The major aim of the study is to develop a novel customer churn prediction model for Sri Lankan Telecommunication Company by considering some soft factors for early identification of customers who leave the service provider. Three machine learning algorithms namely Logistic Regression, Naive Bayes and Decision Tree are used in this study. In fact, twenty attributes are mainly carried out to train these three algorithms. Furthermore, the Back Propagation Neural Network (BPNN) was trained to predict customer churn. In Artificial Neural Network (ANN) training; result of Logistic Regression, Naive Bayes and Decision Tree and eight attributes that mostly affecting the final result are used as inputs. The performances of the models are evaluated by using the confusion matrix using three different data samples. Final ANN model gives 96.7% accuracy in the testing process. Also it gives a high accuracy when comparing with the other data mining algorithms. Existing customer churn prediction models are designed using single algorithm. But the experimental results in this study show multiple algorithms for churn prediction that give higher performance than a single algorithm.

Keywords- machine learning; neural network; algorithm

I. INTRODUCTION

In the past few years, the telecommunication industry in Sri Lanka has shown significant growth compared with other industries. The focus of telecommunication companies has therefore shifted from building a large customer base into keeping customers in house.

Therefore, it is valuable to understand which customers may turn to competitors in the near future. Customer churn prediction was needed to identify these changes of customers. Customers are the major fractions to be focused in every industry as products and services are rendered to them. In fact, efficient business practitioners should be able to cater to the demands of business clients. For that customer churn prediction is very important in any industry. This is same to mobile telecommunication industry too. Because attracting new customers is costlier than protecting their existing customers. For that reason, companies always try to protect their existing customers. To protect their customers, they want to identify customers who have most risk to leave the service.

In this study, customers are post-paid users in Sri Lankan telecommunication Service Company. Post-paid telecom service providers can easily understand churn rate because they work with contracts. When a customer terminates their contract, the company knows they lost a customer. Customer churn occurs when customers or subscribers stop doing business with a company or service provider. It means that they leave the existing company. Customer churning is directly related to customer satisfaction.

There is an intense competition in telecommunication market resulting to introduce more sophisticated products and services. The income can be considerably affected if the company loses its faithful customers. This leads to reduced customer loyalty. Losing an existing high-volume customer means losing lots of revenue. Analysing customer data and customer behaviour is the basis for understanding the needs of any customers. It is necessary to identify customers who are willing to move to a competitor before they do so. The present Sri Lankan mobile industry is extremely dynamic, with new services, technologies, and carriers constantly altering the landscape. Then customers have more choices.

Mobile telecommunication industry generates a huge amount of data like billing information, call detail data and network data. This voluminous amount of data ensures the necessity for the application of data mining techniques in telecommunication database. In the information generated in the telecommunications industry there are hidden data and patterns not yet identified. Analysing this huge amount of data in various perspectives allows service providers to improve their business in various ways.

The major aim of the study is to develop a customer churn prediction model by considering some soft factors like monthly bill, billing complaints, promotions, hotline call time, arcade visit time, negative ratings sent, positive ratings sent, complaint resolve duration, total complaints, and coverage related complaints.

ANN based approaches are mostly used in churn prediction in many subscribers based industries. This study uses 3 machine learning algorithms namely Logistic Regression, Naïve Bayes and Decision Tree with BPNN.

Rest of these is structured as follows. Section 2 discusses the methodology part and describes the technologies used to develop this model. Section 3 describes experimental design. Section 4 describes the result and discussion of the study. This section shows the final result of the model, the prediction accuracy and other performance measurements. Finally, in section 5, the conclusions are given.

II. METHODOLOGY

Before use the dataset for analysing it is necessary to remove incomplete, noisy or inconsistent values in the dataset. In this study WEKA data mining tool was used for data pre-processing.

There are three machine learning algorithms namely Logistic Regression, Naive Bayes and Decision Tree and a BPNN were used in this model. Logistic Regression is a Machine Learning classification algorithm. Logistic Regression analyses studies the association between a categorical dependent variable and a set of independent (explanatory) variables (Arumawadu et al., 2015; Arumawadu et al., 2016). Naive Bayes is one of the speediest statistical classifier algorithm works on probability of all attribute contained in data sample individually and then classifies them accurately. It is used to predict class membership probabilities (Rathnayaka et al., 2012). Decision Tree is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees. Multilayer Perceptron is a nonlinear classifier based on the Perceptron. A Multilayer Perceptron (MLP) is a back propagation neural network with one or more layers between input and output layer.

Logistic Regression, Naive Bayes and Decision Tree algorithms are trained by applying 20 attributes until the output result of each algorithm is matched with the original resultant with minimum error rate for 3 sample datasets and then the training process is stopped. Once the algorithms are trained using these 3 data samples, algorithms with best performance showed dataset was saved.

An Artificial Neural Network with 11 inputs which are 8 input attributes that affect the customer churn mostly and three special inputs, which are prediction result from Logistic Regression, prediction result from Naive Bayes and prediction result from Decision Tree are used in the study. This neural network was trained and best performance model was selected as the final model. Training process was done by adjusting number of neurons in hidden layer1, number of neurons in hidden layer2 and value for the epoch.

Then test the model using testing datasets and the model which gives high performance was saved for use in future predictions. Performance measuring of the models is done

by using some of the confusion matrix-based measures like accuracy, precision, recall or sensitivity and F1 score. As well as used receiver operating characteristic (ROC) analysis.

Data set used in this study contains 3,334 subscribers, including 1,289 churners and 2,045 non-churners. Three data samples were used to train and test the model. Such as a data sample with 60% for training and 40% for testing, a data sample with 70% for training and 30% for testing and a data sample with 80% for training and 20% for testing.

III. EXPERIMENTAL DESIGN

When using neural networks to perform predictive modelling, the input layer contains all of the input fields or variables used to predict the outcome variable. The output layer contains an output field which is the target of the prediction. The input and output fields can be numeric or symbolic.

ANN was trained by changing number of hidden neurons and epoch until achieving better prediction accuracy. In this study first hidden layer neurons changed from one to fifteen. Second layer neurons changed from one to ten. Epoch was changed from 10 to 40. All these tests were done for all datasets such as a data sample with 60% for training and 40% for testing, a data sample with 70% for training and 30% for testing and a data sample with 80% for training and 20% for testing.

IV. RESULTS AND DISCUSSION

First the logistic regression algorithm was trained and tested using 3 data samples. After training and testing the logistic regression algorithm using these 3 data sets, the model that showed the best accuracy was selected to apply in the final prediction model. Table 1 shows the training and testing accuracy of the logistic regression algorithm.

Table 1. Training and testing accuracy score of the logistic regression algorithm.

	Training	Testing
Accuracy (60/40)	0.8689	0.8564
Accuracy (70/30)	0.8767	0.8622
Accuracy (80/20)	0.8794	0.8525

Then Naïve Bayes algorithm was trained and tested using 3 data samples and the model that showed was saved to use final prediction model.

Table 2. Training and testing accuracy score of the Naïve Bayes algorithm.

	Training	Testing
Accuracy (60/40)	0.9218	0.9058
Accuracy (70/30)	0.8565	0.9333
Accuracy (80/20)	0.9187	0.9203

Next the Decision Tree algorithm was trained and tested using 3 data samples. After training and testing the Decision Tree algorithm using these 3 data sets best accuracy shown model was selected to apply in final prediction model.

Table 3. Training and testing accuracy score of the Decision Tree algorithm.

	Training	Testing
Accuracy (60/40)	0.9218	0.9058
Accuracy (70/30)	0.8565	0.9333
Accuracy (80/20)	0.9187	0.9203

A modified ANN was built to perform customer churn prediction. Inputs for final model are 8 attributes most affect for churn and result of Logistic Regression, Naïve Bayes and Decision Tree algorithm. Trained ANN with 11 inputs in input layer, 2 hidden layers and one output layer is used in this model.

Result of the study is a modified Artificial Neural Network with one input layer, two hidden layers and one output layer gives 96.7% accuracy score for mobile telecommunication customer churn prediction. Performance of the model was shown in table 4.

Table 4. Performance of the novel customer churn prediction model

	For Training data set	For Testing data set
Accuracy Score	0.9863	0.9677
Precision Score	0.9804	0.9788
Recall Score	0.9946	0.9820
f1 Score	0.9875	0.9804

The structure of the constructed ANN is illustrated in Figure 1. It consists 4 layers which are input layer with 11 input neurons, hidden layer 1 with 15 input neurons, hidden layer 2 with 10 input neurons and output layer with neuron. A training epoch is set to 40.

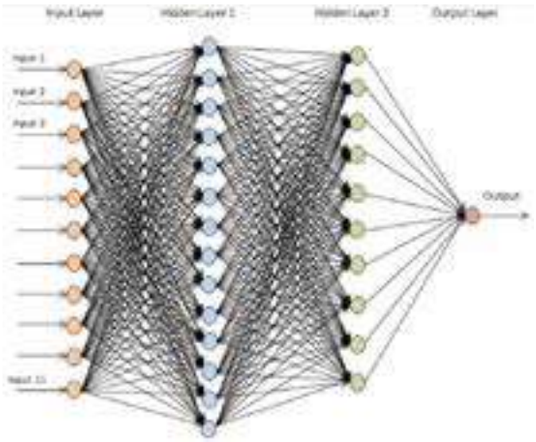


Figure 1. Structure of the constructed ANN model

V. CONCLUSION

A first logistic regression algorithm was trained using all 20 attributes. That trained model shows good performances. Accuracy score for this model is 84.7%. Then naive bayes algorithm was trained using same attributes. This model shows 91% accuracy in testing data set. Finally, decision tree algorithm was trained using same attributes. It shows 81.74% accuracy for testing data set.

Then novel model was trained using result of the above three algorithms and 8 attributes most affect for final result. The findings of this study confirm that churn can be predicted successfully with certain level of accuracy using this novel model. Developed model achieve above 95% overall accuracy on a testing set.

ANN with other three machine learning algorithms creates a strong customer churn prediction model. Performance of novel model is higher than using them separately.

REFERENCES

Almana A, Aksoy M, Alzahrani R, (2014) A survey on data mining techniques in customer churn analysis for telecom industry, *International Journal of Engineering Research and Applications*, 45, 165-171.

Eriksson K, Vaghult A, (2000) Effects on New Customer Acquisition by Retained Customers in Professional Services, In *16th IMP Conference*, Bath.

Arumawadu HI, Rathnayaka RMKT, Seneviratna K, (2016) New Proposed Mobile Telecommunication Customer Call Center Roster Scheduling Under the Graph Coloring Approach, *International Journal of Computer Applications Technology and Research*, 5(4), 234–237.

Arumawadu HI, Rathnayaka RMKT, Illangarathne SK, (2015) K-Means Clustering For Segment Web Search Results, *International Journal of Engineering Works*, 2(8), 79-83.

Arumawadu HI, Rathnayaka RMKT, Illangarathne SK, (2015) Mining Profitability of Telecommunication Customers Using K-Means Clustering, *Journal of Data Analysis and Information Processing*, 3(3), 63.

Rathnayaka RMKT, Zhong-jun W, (2012) Enhanced Greedy Optimization Algorithm with Data Warehousing for Automated Nurse Scheduling System, *E-Health Telecommunication Systems and Networks*, 1(4).

Hung S, Yen D, Wang H, (2006) Applying data mining to telecom churn management, *Expert Systems with Applications*, 31(3), 515-524.

Kirui C, Hong L, Cheruiyot W, Kirui H, (2013) Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining, *IJCSI International Journal of Computer Science Issues*, 10(2), 1694-0814.

Markopoulos A, Manolakos D, Vaxevanidis N, (2008) Artificial neural network models for the prediction of surface roughness in electrical discharge machining. *Journal of Intelligent Manufacturing*, 19(3), 283-292.

Sweeney J, Swait J, (2008) The effects of brand credibility on customer loyalty, *Journal of retailing and consumer services*, 15(3), 179-193.

IMPACT OF BIG DATA AND POLITICAL MICROTARGETING ON DONALD TRUMP'S 2016 PRESIDENTIAL CAMPAIGN

Narishia Singh[1] and Anuradha Wijegunawardhana[2]
[1]nsingh@nrcocp.com, USA
[2]anuradhawije@nrcocp.com, Sri Lanka

Abstract- On January 20th, 2017, Donald Trump was inaugurated as the 45th President of the United States. His opponent, Hillary Clinton, was a seasoned politician who was favored to win the election. However, Trump produced one of the biggest political upsets in modern day history. This explanatory case study aimed to explain one of the reasons that Trump won by looking at the role that technology played during the 2016 presidential election. Specifically, this study examined how big data was leveraged to influence voter behavior during the campaign. Data drawn from articles, interviews, and videos published during the campaign were collected and analyzed thematically. Results identified three main themes that were particularly salient: mobilization of voter databases, data-driven microtargeting, and Facebook influence. This study highlightes the potential benefits of targeting voters using big data analytics as well as the potential risks related to issues of privacy.

Keywords- Big Data, political microtargeting, psychographics, election campaigns

I. INTRODUCTION

The proliferation of technology in society has had a major impact on political campaigns. The field of big data specifically played a crucial role in the recent 2016 US presidential election with the rise of data-driven political microtargeting fuelled by the Trump campaign. Some analysts note that Trump's victory was based on his campaign's ability to target specific segments of the electorate more effectively than Clinton. This explanatory case study thereby aimed to understand how big data and political microtargeting were used during Trump's 2016 presidential campaign. The following question guided the study: 1) How did the Trump campaign leverage big data to influence voter behaviour?

Prior research that examined how political candidates in the U.S. gather and manipulate data to appeal to voters, has focused on the Obama campaigns mainly (Tufekci, 2014). This case study will, therefore, contribute to research in this field.

II. LITERATURE REVIEW

Big data election campaigns have gained increased momentum in the last decade. Access to a huge volume and variety of voter data, along with a growing infrastructure of cutting-edge technology tools now allow campaign teams to gain deep insight about each constituent. Although the term itself has numerous definitions, for the context of this paper, big data is defined as "large pools of data that can be captured, communicated, aggregated, stored, and analyzed" (McKinsey Global Institute, 2011, p. 1). Analytics is a key aspect of big data since it allows data-driven decisions to be made. "To help decision making, data analysts choose informative metrics that can be computed from available data with the necessary algorithms and tools and report the results in a way the decision makers can comprehend and act upon" (Lau, Yang-Turner and Karacapilidis, 2014, p. 50). Candidates now rely heavily on data collection and analytics to shape their communication strategies, specifically as it relates to political microtargeting. "Through political microtargeting, a political party can identify the individual voters which it is most likely to convince" (Borgesius et al., 2018, p. 82). They can then use various campaign appeals (e.g. door-to-door conversations, direct mail, social media messages, emails or texts) to connect with these voters using persuasive messages.

Research about big data and political microtargeting can be grouped into three broad topical categories: voter data, political data analytics, and voter microtargeting. A brief overview of each category is discussed below.

1. Voter Data: The United States has maintained electronic voter profiles with data points for each registered voter, for over two decades. Research by Alexander and Mills (2004) show that state voter registration databases usually contain the name, address, date of birth, and phone number of each individual along with past electoral participation. On a national level, both the Republicans and Democratic parties maintain expansive digital databases (known as voter files) that merge the states' voter data with other forms of public data (e.g. motor vehicle, real estate and campaign donation records) and consumer data (e.g. credit reports) as well as campaign web data and data acquired from data brokers (Howard and Kreiss, 2010; Rubinstein, 2013). For example, as of 2015 the Republican party's GOP Data Center contained information on over "200 million individuals with over 700 billion data points" (GOP, 2015). An increasing number of partisan and nonpartisan political firms also offer commercial voter files. As an example, long-standing non-partisan firm Aristotle maintains a national voter file with "over 190 million records that are enhanced with phones, emails, demographics and lifestyle information" as well as a separate consumer database and donation file (Aristotle). Akosah (2015) notes that the ongoing rise of social media and digital devices means that voter files are constantly updated with new information from online sources, including campaign website registration and online tracking cookies. Voter data is also accumulated from offline sources such as volunteer and fundraising events. (Akosah, 2015).

2. Political data analytics: Campaigns analyze voter data "to form predictions about which members of the public are supporters, which are likely to show up to vote, and which are persuadable" (Hersh, 2015, p. 3). Based on these predictions, campaigns can make strategic decisions about who to target and formulate fundraising objectives. Nickerson (2014, p.58) notes that by using voter data, "campaigns are able to predict with greater accuracy which citizens will support their candidates and issues better than which citizens will oppose [them]." For example, by looking at the attributes of citizens who donate or volunteer, campaigns can design models to identify correlations among similar citizens (Nickerson,2014). Data analysts use sophisticated statistical techniques to create predictive models from a universe of voters; such as a contactable, support, turnout and persuasion model.

Figure 1. Creating Persuasion Universes in Battleground States Source: Fulgoni, Lipsman and Davidsen, 2016

For example, during his re-election campaign in 2012, Barack Obama's campaign created a "likelihood to turnout" model using surveys and the vast datasets collected about each voter (Tufekci, 2014). After assigning a score from 0-100 to determine a voter's likelihood of voting, the campaign used this information to target individuals who scored high on the index (Tufekci, 2014).

3. Voter Microtargeting: Vast sets of voter data allow campaigns to target individual constituents with specific messages. For example, instead of targeting broad demographic communities (e.g. women or minorities), campaigns can segment the population and appeal to a conservative white male living in Jacksonville, Florida or a liberal African-American living in Philadelphia. According to Howard and Erickson (2009), this benefits a campaign since they can allocate valuable funds towards voters who are receptive to their message rather than focusing on unlikely voters. Microtargeting works by first collecting all of the aggregated information in each voter's profile including survey, "demographic, geographic and marketing data…" (Bennett, 2015, p.374). Predictive models are then created and "hundreds of customized messages for each [constituent], each with highly personalized political content…" are designed and then delivered using various channels of communication (International Institute for Democracy and Electoral Assistance, 2018, pp..13-14). Recent research conducted by Schipper and Woo (2017, pp. 18, 22) shows that microtargeting is necessary for a voter to view a candidate favorably since individuals who do not

Table 1. How big data has changed traditional targeting into microtargeting
Source: International IDEA

Traditional targeting	Digital microtargeting
Collecting data	Increased availability of big datasets: collected by parties themselves, government agencies, polling agencies, voter files, as well as consumer data purchased from commercial market research firms Data can be collected more easily: citizens' personal information can be reached more readily online, as can their digital footprint Data can be stored more easily through larger servers. For example, US President Donald Trump's election campaign had 'more than 300 terabytes of data' (Halpern 2017)
Dividing voters into segments based on characteristics such as personality traits, interests, background, or previous voting behaviour	'Predictive analytics': patterns can be recognized more easily with the use of complex algorithms 'Psychological targeting': squaring voter data collected by political parties with consumer data purchased from commercial market research firms; this helps to build a more detailed profile: what people buy, eat or watch in some cases can help to predict how they vote. The impact of psychological targeting is being debated.
Designing personalized political content for each segment	'A/B testing': sending out hundreds of thousands of slightly different versions of the same message to different population segments to test patterns in their responses, such as how quickly they click, how long they stay on a page, what font and colour layout they like
Using communication channels to reach the targeted voter segment with tailor-made messages	Pairing voter profiles with social media user data to reach the right people with the right message

receive targeting messages will be less likely to infer vague policy points from a candidate.

Segmenting the population using psychographics is one way to send individualized messages to voters. This data includes information about a voter's personality, beliefs, values, attitudes, and opinions. Prior research indicates that individuals are more receptive to messages that are specifically tailored to match their preferences. For example, in a study where 3.5 million people were exposed to psychologically tailored advertising in three experiments, results showed that people bought or "liked" more products when they were shown online ads tailored to match their personality (Matz, Nave and Stillwell, 2017).

III. METHODOLOGY

This study followed a qualitative method of research using an explanatory, single-case study. In defining what a case study is, Yin suggests that the term refers to an event, an entity, an individual or even a unit of analysis. It is an empirical inquiry that investigates a contemporary phenomenon within its real-life context using multiple sources of evidence (Yin, 1989). An explanatory case study was selected as the preferred strategy since it specifically aims to answer "how" or "why" questions in-depth, using multiple sources.

Data collection consisted of an eleven-minute online video of Alexander Nix, the former CEO of Cambridge Analytica who gave an overview of big data and psychographics and highlighted the company's collaboration with Senator Ted Cruz's presidential campaign. Similarly, a 60 Minutes interview featuring Trump's digital director Brad Parscale was included. In addition, an online article published by the magazine Businessweek was also sourced since it focused on the Trump campaign's real-time data operation twelve days before the presidential election. Two thirty-second advertisements that the campaign aired were also included. To add further insight about this research, a dataset of 24 online documents related to big data and political microtargeting was sourced using the Google Scholar search engine. To address the issue of validity and reliability, triangulation of the data was entailed; documents were drawn from various disciplines including communication, economics, law, marketing and political science.

In terms of analytical strategies, thematic analysis was selected since it provides a systematic yet flexible element to data analysis. As Braun and Clarke (2013, p. 6) note, "thematic analysis is essentially a method for identifying and analyzing patterns in qualitative data." The goal is to identify themes from a broad set of data by determining the relationships between concepts. Braun and Clarke's iterative six-step framework is widely considered the most influential approach for doing thematic analysis and was therefore utilized in this study.

IV. RESULTS

Results show that there were three ways the Trump campaign leveraged big data to effectively target key voters.

1). Mobilization of Voter Databases: The Trump campaign accessed a trove of voter data by utilizing three voter databases during the election. First, was the deployment of a custom-designed fundraising database named Project Alamo which was initiated in November 2015. The database quickly expanded after Trump became the Republican nominee for President in July 2016. Data was gathered from small donor contributions, email addresses gathered at rallies, sales of campaign merchandise and text messages sent to the campaign (Halpern, 2017). Likewise, Cambridge Analytica also designed a specialized database for the campaign; however, it was based largely on psychographic information about America voters. In a presentation given at an annual non-partisan summit in September 2016, Alexander Nix described how his political consulting firm administered a Big Five personality survey (they gave it the acronym OCEAN) to hundreds and hundreds of thousands of Americans (Concordia, 2016). Along with the survey, the company also acquired data about individuals using sources such as "…digital data, voter history, and marketing resources supplied by leading companies, including Acxiom, Experian, Nielsen…and Facebook…" (Chester and Montgomery, 2017). Cambridge Analytica claimed that all this data was then aggregated using big data analytics to create a personality profile for every adult in the United States.

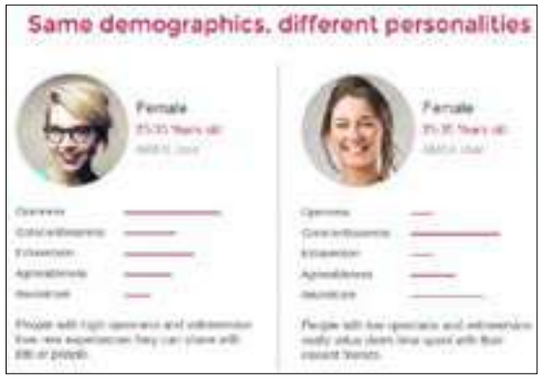


Figure 2. Same demographics, different personalities
Source: Computer Business Review

Finally, the campaign relied on the GOP Data Center, the 20-year old national voter file. Eventually, Cambridge Analytica's database and the GOP Data Center would be merged with Project Alamo, to create a massive database that held the profiles of over 220 Americans. Each profile "contained between 4,000 and 5,000 data points...of individual voters" (Owen, 2018, p. 41). `

2). Data-Driven Microtargeting: Using the data gathered on the electorate, the Trump campaign micro-targeted precise segments of the population. For example, data from the GOP Data Center revealed that infrastructure was an issue of concern to some voters. In a 2017 profile on 60 minutes, the campaign's digital director Brad Parscale discussed how "voters in the Rust Belt cared about their roads being built, their highways, their bridges" (CBS News). He then started making ads that would address these issues to target persuadable voters in states such as Pennsylvania and Michigan. The campaign continued to focus on the Rust Belt states as the campaign progressed.

In October 2016, the campaign's data analytics team (which consisted of four data firms, including Cambridge Analytica) used regularly-updated voter data to identify three segments of the population where Trump needed to win: unallocated voters who could be persuaded by Trump's message, unconvinced Republicans, and voters who were supporting Hillary Clinton but wanted changes in government (Kaye, 2016). The campaign then ran television ads to appeal to these voters. For example, one advertisement targeted male voters supporting Clinton; it ran during a pickup truck racing series in Toledo, Ohio and portrayed Mr. Trump as a strong leader who would fight for American workers (Deals). Another advertisement focused on women who were Clinton supporters; it ran during a daytime talk show in states such as Wisconsin and focused on "Trump's promises to provide childcare tax reduction and paid maternity leave" (Builders).

As the campaign entered its final stages, Trump's digital team microtargeted voters on a much larger scale. It ran models which showed three sectors of the population where Hillary Clinton needed to capture votes and then devised a plan to discourage these groups from voting. In Green and Issenberg's (2016) exclusive Businessweek article (which gave empirical evidence about the Trump campaign), a senior official discussed "three major voter suppression operations under way" that targeted idealistic white liberals, young women, and African Americans. To depress voter turnout among African Americans specifically, the campaign created a cartoon animation in which Hillary Clinton called black gang members "super predators." They then used the ads in Facebook "dark posts" which were non-public, paid posts that were only shown to selected Facebook users (Green and Issenberg, 2016).

Likewise, the decision to canvass in Michigan (the heartland of America's automotive industry) during the final weeks of the campaign was based on Cambridge Analytica algorithms and psychographic data about the state's voters. Once the company divided the US population into 32 personality types and focused on 17 states, the campaign discovered that "a preference for cars made in the US was a great indication of a potential Trump voter" (Grassegger and Krogerus, 2017). As a result, the Trump campaign made last minute campaign stops in the Rust Belt state before the election and consequently won Michigan (along with Wisconsin and Pennsylvania) to secure an electoral college win.

3). Influence of Facebook: Facebook played a key role in helping the Trump campaign convey its messages to voters. In his interview, Parscale claimed that "he understood early that Facebook was how Donald Trump was going to win" and he spent much of his budget on Facebook ads (CBS news). For example, in the early stages of the primaries, Parscale launched microtargeted advertisements on Facebook using data collected about all known Trump supporters. The advertisements relied heavily on Facebook's powerful custom audience targeting features which allowed the campaign to target voters based on interests and behaviors (Halpern, 2017). From there, Parscale expanded his pool of targets using "…Facebook's Lookalike Audiences tool to find people with interests and qualities similar to those of his original cohort and developed ads based on these characteristics…" (Halpern, 2017). During the third presidential debate, the campaign also tested 175,000 different variants of an advertising post on Facebook. "The messages differed for the most part only in microscopic details, in order to target the recipients in the optimal psychological way: different headings, colors, captions, with a photo or video" (Grassegger and Krogerus, 2017).

V. DISCUSSION

This study examined how big data and political microtargeting were used during Donald Trump's 2016 presidential election campaign. As the results show, the widespread availability of huge sets of voter data made it possible for the campaign to gain deep insight about each voter during the election. The campaign then used this information to create adverts that targeted specific individuals. The use of psychographics added an advantage; as Alexander Nix, the former CEO of Cambridge Analytica noted, "If you know the personality of the people you are targeting, you can nuance your messaging to resonate more effectively with those key audience groups" (Concordia, 2016). Finally, Facebook offered a broad spectrum of commercial digital marketing tools and techniques which the campaign used to appeal to voters.

From a broader perspective, several key observations can be made from the research. First, Trump's 2016 campaign shows that big data now plays a critical role in politics. Campaigns have access to a vast amount of voter data which they can analyze to create a comprehensive profile of each member of the electorate. Elections are also becoming more localized; as campaigns shift away from targeting voters across broad communities. At the same time, social media continues to be a main communication channel between candidates and voters. Nevertheless, this study raises several important issues about data privacy. Concerns have been raised about the way Cambridge Analytica accessed voter data and the company dissolved earlier this year as media reports indicated that the company allegedly received psychographic data from millions of Facebook users without their consent.

Questions about ethics also arise because of the Cambridge Analytica scandal. Big data allows for highly personal information to be mined about people on a massive scale, yet privacy regulation in the United States is very weak. Individuals, therefore, have no significant data protection when their data is breached or used without consent. Having access to people's digital footprints also creates a threat of manipulation with political microtargeting. "For instance, a party could target particular voters with tailored information that maximizes, or minimizes, voter engagement" (Borgesius et al., 2018). One, therefore, must wonder whether politicians will use the valuable insight that big data provides to make unfair or discriminatory decisions during an election campaign.

VI. CONCLUSION

The Trump presidential campaign of 2016 revealed how big data and political microtargeting are transforming political elections. Through an analysis of data sources, this case study revealed that big data was used to tactically target specific individuals and make large-scale strategic decisions. From a broader perspective, this study demonstrates how a mass analysis of individual data can have a strong impact on both political communication and the political system itself. Nevertheless, this study has several limitations. Campaigns driven by big data are a new phenomenon and although Important journalistic commentary exists about the topic, academic research is fragmented. For example, there is very little research available about the methodology of mining voter data to discover correlations and patterns. Likewise, as a single case study, this research cannot be generalized to a wider population. Further research that investigates how big data and political microtargeting are used in elections will, therefore, be fruitful.

REFERENCES

Alexander, K. and Mills, K. (2004). *Findings from the California Voter Foundation's 2002 state-by-state survey of voter registration data gathering and privacy practices*. [online] Davis:California Voter Foundation, p. 3. Available at: https://www.calvoter.org/issues/votprivacy/pub/0504voterprivacy.pdf [Accessed 13 July. 2018].

Akosah, K.N. (2015). Cracking the One-Way Mirror: How Computational Politics Harms Voter Privacy, and Proposed Regulatory Solutions. *Fordham Intellectual Property, Media and Entertainment Law Journal* [online] Volume 25 (4), p. 1019. Available at: https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com/&httpsredir=1&article=1603&context=iplj [Accessed 14 July. 2018].

Aristotle. Campaigns Features. [online]. Available at: http://aristotle.com/campaigns/features/ [Accessed 16 July. 2018].

Bennett, C.J. (2015). Trends in Voter Surveillance n Western Societies: Privacy Intrusions and Democratic Implications. Surveillance & Society, [online], Volume 13 (3/4), p. 374. Available at: https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/voter_surv/voters [Accessed 14 July. 2018].

Borgesius, F.Z, Moller, J., Kruikemeier, S., Fathaigh, R.O., Irion, K., Dobber, T., Bodo, B. and Vreese, C. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*. [online] Volume 14 (1), p. 82. Available at https://www.utrechtlawreview.org/articles/abstract/10.18352/ulr.420/ [Accessed 03 May. 2018].

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, [online] Volume 3 (2), p. 6. Available at http://eprints.uwe.ac.uk/11735/2/thematic_analysis_revised... [Accessed 10 May. 2018].

Builders. [video]. Available at http://adage.com/article/campaign-trail/trump-camp-s-inexperience-set-stage-rnc-data-win/307105/ [Accessed 16 July. 2018].

CBS News. (2017). Secret Weapon. [video]. Available at https://www.cbsnews.com/video/secret-weapon/ [Accessed 09 May. 2018].

Chester, J. and Montgomery, K.C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*. [online] Volume 6 (4), p. 7. Available at https://policyreview.info/articles/analysis/role-digital-marketing-political-campaigns [Accessed 02 May. 2018].

Computer Business Review. Same demographics, different personalities. [image]. Available at:https://www.cbronline.com/news/big-data/analytics/british-big-data-win-trump-u-s-election/attachment/cambridge-analytica/ [Accessed 16 July. 2018].

Concordia. (2016). *Cambridge Analytica – The Power of Big Data and Psychographics*. [video]. Available at https://www.youtube.com/watch?v=n8Dd5aVXLCc [Accessed 01 May. 2018].

Deals. [video]. Available at http://adage.com/article/campaign-trail/trump-camp-s-inexperience-set-stage-rnc-data-win/307105/ [Accessed 16 July. 2018].

Fulgoni, G.M., Lipsman, A. and Davidsen, C. (2016). *Creating Persuasion Universes in Battleground States.* [image] Available at https://www.researchgate.net/publication/308274079_The_Power_of_Political_Advertising_Lessons_for_Practitioners_How_Data_Analytics_Social_Media_and_Creative_Strategies_Shape_US_Presidential_Election_Campaigns [Accessed 15 July. 2018].

GOP. (2015). *RNC Launches Data Center 2016.* [online] Available at https: // www .gop. com / rnc - launches - data - center - 2016 / [Accessed 13 July 2018].

Grassegger, H. and Krogerus, M. (2017). *The Data That Turned The World Upside Down*. [online] Stanford Public Policy Program. Available at https://publicpolicy.stanford.edu/news/data-turned-world-upside-down [Accessed 05 May. 2018].

Green, J. and Issenberg, S. (2016). *Inside the Trump Bunker With Days to Go.* [online] Bloomberg Businessweek. Available at https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go [Accessed 02 May 2018].

Halpern, S. (2017). *How He Used Facebook to Win*. [online] The New York Review of Books. Available at http://www.nybooks.com/articles/2017/06/08/how-trump-used-facebook-to-win/ [Accessed 07 May. 2018].

Hersh, E. (2015). *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge University Press, p. 3.

Howard, P. and Erickson, K. (2009). Data Collection and Leakage. *Chicago-Kent Law Review*, [online], p. 738. Available at https: // scholarship. kentlaw. iit. edu / cgi / view content. cgi?article=3713&context=cklawreview [Accessed 12 July. 2018].

Howard, P. N. and Kreiss, D. (2009). *Political Parties & Voter Privacy: Australia, Canada, the United Kingdom, and United States in Comparative Perspective*. [pdf]. Seattle: University of Washington, p. 21. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2595120 [Accessed 01 July. 2018].

International Institute for Democracy and Electoral Assistance. (2018). *Digital Microtargeting: Political Party Innovation Primer 1*. [online] Stockholm: International IDEA, pp. 13-14. Available at https: // www. idea. int /sites / default / files / publications / digital-microtargeting.pdf [Accessed 15 July. 2018].

Kaye, K. (2016). How The Trump Camp's Data Inexperience Helped Propel His Win. [online] AdAge. Available at http://adage.com/article/campaign-trail/trump-camp-s-inexperience-set-stage-rnc-data-win/307105/ [Accessed 16 June. 2018].

Lau, L., Yang-Turner, F. and Karacapilidis, N. (2014). Requirements for Big Data Analytics Supporting Decision Making: A Sensemaking Perspective. In: N. Karacapilidis, ed., *Mastering Data-intensive Collaboration and Decision Making,* 1st ed. [ebook] Springer International Publishing, p. 50. Available at https: / /pdfs. semanticscholar. org / d996/d55946be27c73d1ff05f98e0b44d6e51ccf8.pdf?_ga=2.132493452.347085874.1532377787-1997264.1532377787 [Accessed 08 May. 2018].

Matz, S.C., Nave, G. and Stillwell, D.J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America,* [online] Volume 1114 (48), p. 12714. Available at http : // www. pnas. org / content / pnas / 114 /48/ 12714. full. pdf [Accessed 10 May. 2018].

McKinsey Global Institute. (2011). Big data: *The next frontier for innovation, competition, and productivity*. [pdf] McKinsey and Company, p. 1. Available at https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx [Accessed 12 July 2018].

Nickerson, D.W. and Rogers, T. (2014). Political Campaigns and Big Data. *Journal of Economic Perspectives*, [online] Volume 28 (2), p. 58. Available at: https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.51 [Accessed 09 May. 2018].

Owen, D. (2018). Characteristics of US Elections in the Digital Media Age. In: S. Kiyohara, K. Maeshima and D. Owen, ed., Internet Election Campaigns in the United States, Japan, South Korea, and Taiwan, 1st ed. [ebook] Palgrave Macmillan, p. 41. Available at http://lib.bvu.edu.vn/bitstream/TVDHBRVT/18387/1/978-3-319-63682-5.pdf#page=41 [Accessed 15 July. 2018].

Rubinstein, I.S. (2014). Voter Privacy in the Age of Big Data. *Wisconsin Law Review*, [online] p. 879. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2447956## [Accessed 13 July. 2018].

Schipper, B.C. and Woo, H. (2017). *Political Awareness, Microtargeting of Voters, and Negative Electoral Campaigning*. [online]. SSRN, pp. 18,22. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039122 [Accessed 15 July. 2018].

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, [online] Volume 19 (7). Available at http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097 [Accessed 15 July. 2018].

Yin, R.K. (1984). *Case Study Research: Design and Methods*. California: Sage, p. 23.

PREDICTING THE RISK OF BEING A DIABETIC PATIENT USING STATISTICAL ANALYSIS AND DATA MINING

BPN Perera[1] , BTGS Kumara[2], and HACS Hapuarachchi[3]

[1,2]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka

[3]Department of Sports Science and Physical Education, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka

[1]*pinipanisansala359@gmail.com*

Abstract-There is a vast and enormous amount of data available in hospitals and medical related institutions. But, the amount of knowledge obtained from such data is very little. Applying IT knowledge for healthcare is an emerging field of huge importance for providing prognosis as well as a deeper understanding of medical data. Diabetics is actually a disease which is affecting many people today Early prediction of diabetes is an extremely challenging task because of the complicated interrelationship between various factors. This research tried to diagnose diabetes which based on 12 risk factors using data from 200 people and applied data mining and statistical techniques to predict the risk of being a diabetic patient. Statistical model has been created using Minitab with the application of the binary logistic regression model. The created model provided the way of predicting the possibility of having diabetics for any person and identified the most suitable risk factors which are most relevant to the disease prediction. Through this identified risk factors, we clustered the data using k-means. An empirical study has proved the effectiveness of our proposed approach.

Keywords- Binary logistic regression, Data mining, K-means clustering

I. INTRODUCTION

Diabetes is a disease that occurs when the blood glucose, often known as blood sugar, is simply too high. Blood glucose is the prime source of strength and is derived with the foodstuff people eat. Further, diabetes is a disease that occurs just as the insulin manufacturing inside the body is insufficient or the body is not able to use the produced insulin in a correct manner, consequently, the aforementioned one leads to high blood glucose. Insulin, a hormone produced by the pancreas, helps sugar originating at food deal with the cells for use for strength and energy. Sometimes the body does not conduct enough or any insulin or does not use insulin properly. The body cells dismantle the food into glucose and this glucose should be shifted to all of the cells of the body. After that prevail person blood and does not reach his/her cells. Another word, the insulin is definitely the hormone that directs the glucose which is produced by breaking off the food into the body cells. Any change within the manufacturing of insulin results in an increase in the blood sugar levels and this can lead to damage to the tissues and failure of the organs (Kelly, 2006).

There are three main types of diabetes:

- Type 1: It occurs when the body doesn't manufacture insulin. The immune system of the body attacks and destroys the cells in the pancreas which make insulin. Though there are actually only around 10% of diabetes patients experience this type of Diabetes. The disease reveal as an autoimmune disease happening at a very young age of less than twenty years thus further called juvenile-onset diabetes. It is often diagnosed in teenagers and youthful adults, although it can present at any age.
- Type 2: It occurs when the body doesn't manufacture or use insulin properly. Type 2 is the commonest type of diabetes. It can happen and develop type 2 diabetes at any age, even all through childhood. This type accounts for nearly 90% of the diabetes cases

and generally known as the adult-onset diabetes or the non-insulin dependent diabetes. In this case, the several organs of the body turn into insulin resistant, and that increases the demand for insulin. At that case, pancreas doesn't manufacture the necessary amount of insulin. To keep this kind of diabetes at bay, the patients have to follow a strict diet, exercise routine and keep track of the blood glucose. Obesity, being overweight, being physically inactive can lead to type 2 diabetes.

- Gestational diabetes pursues some females when they are pregnant. Most of the time, this type of diabetes leave after the child is born. It is actually a type of diabetes that has a tendency to reveal in pregnant women because of the high sugar levels as the pancreas don't produce sufficient amount of insulin. Taking no treatment can result in complications for the duration of childbirth. Controlling the diet and taking insulin can regulate this kind of diabetes. However, if someone had gestational diabetes, then she has got a super chance of coming up type 2 diabetes thereafter in life (Your guide to diabetes : type one and type two, 2013).

Among these three types of diabetic, type 2 diabetic was selected for the research. According to International Diabetes Federation 382 million people are living with diabetes all over the world. By 2035, this would be doubled as 592 million. Diabetes led to 1.5 million deaths in 2012. Higher-than-optimal blood glucose was responsible for an extra 2.2 million deaths due to expanded risks of cardiovascular and different diseases, for a whole of 3.7 million deaths related to blood glucose levels in 2012. Many of these deaths (43%) occur below the age of 70. In 2014, 422 million people in the world had diabetes – a prevalence of 8.5% among the grownup population. The prevalence of diabetes has been unwaveringly increasing for the past 3 decades and is growing most rapidly in low- and middle-income countries. In 2011, 71.4 million people (8.3%) in South East Asia were affected by diabetes and 23.8 million people (2.8%) were affected by Impaired Glucose Tolerance (IGT). Numbers are expected to reach 120.9 million (10.2%) for diabetes and 38.6 million (3.2%) for IGT by 2030, according to the DASL. The diabetes prevalence of the people over two decades in Sri Lanka reveal that the urban population was 16% and among the rural population it was 8%. Under the age of 20 years it was 8.2%. In United States, 30.3 million people (9.4% of the population) had diabetes in 2015 (GLOBAL REPORT ON DIABETES, 2016).

II. RELATED WORKS

(Lowanichchai et al., 2006) proposed an approach using decision tree. The research used data originating at health description, Suranaree Army2 Hospital. It used screening data for diabetes risk which might be used for decision support in making plans the right treatment and relevant for individual patients. The result showed that the Random Tree model has the highest accuracy and the NB Tree model has minimal accuracy. Research paper (Yuvarani & Selvarani, 2016) shows the contrast of the three decision tree classification models for the UCI storehouse diabetes dataset and shows the tree structure formed enabling users to receive truthful decisions in accordance with the input parameters. Further, the genetic J48 model is found to be the most valuable and truthful when compared with any other two decision models in terms of time, efficiency and features.

(Huang et al., 2007) employed three data mining algorithms namely Naive Bayes, IB1 and C4.5 to predict diabetes on data gathered from Ulster Community and Hospitals Trust (UCHT) between 2000 and 2004. They were able to achieve an efficiency of 98%.

Research work (Rajesh & Sangeetha, 2012) has recorded a research study by the use of a variety of classification algorithms like ID3, C4.5, LDA, Naïve Bayes, K-NN for diagnosing diabetes for the given dataset. It has showed which C4.5 is definitely the most competitive algorithm with minor error rate of 0.0938 and further efficiency value of 91%.

(ferreira, oliveira & freitas, 2012) used the several classification algorithms like simplecart, j48, simple logistics, smo, naivebayes as well as bayesnet in order to get diagnosing neonatal jaundice in type1 diabetes. Among all algorithms, it found that Simple Logistics as the finest algorithm. Further, (Parthiban, Rajesh & Srivatsa, 2011) practiced Naïve Bayes approach to determine heart related problems that are occurring in diabetic patients.

However, compare to other research we used two approach to predict and analyse diabetic patients. First, we generated statistical model. Clustering approach is used as second method. In clustering approach, we clustered patients according to the risk level.

III. METHODOLOGY

In this research, Binary logistic regression is used for the creating the statistical model to calculate the possibility of being a diabetic patient. Next, Simple K-means clustering algorithm is used for predicting the level of diabetes.

A. Dataset Used

The data which is used in this research has records of 200 from the government hospital in Kiribathgoda, Sri Lanka. The dataset includes details of both diabetic patients and non-diabetic patients of all age group and Table 1 shows all the attributes used for this research work

Table 1. The attributes used in Data Set for this research

Attribute Name	Values
Age	Real Numbers
Sex	Male or Female
Sugar Consumption	0 or 1
BMI value	Real Numbers
Disease of Pancreas	0 or 1
Diastolic Blood Pressure	Real Numbers
Cholesterol Availability	0 or 1
Background (Family History)	0 or 1
Alcohol Consumption	0 or 1
Smoking	0 or 1
History of Heart Disease/ Stroke	0 or 1
Infection/Illness/skin rashes	0 or 1
No of Children	Real Numbers
Dark, thick skin around neck	0 or 1

B. Data Pre-processing

The Dataset used in this research is medical dataset which may have some inconsistencies. To remove those inconsistencies data pre-processing is done. In Data pre-processing the misclassified data is removed. In this process cleaning and filtering of the data is carried out with respect to the data and data mining algorithm employed so as to avoid the creation of deceptive or inappropriate rules or patterns.

C. Research Process

- Step 1: Applying Binary Logistic Regression to the data set for creating statistical model and identify most effective attributes.
- Step 2: Applying Simple K-Means algorithm to the dataset of most effective attributes for clustering the data into four clusters as normal, impaired fasting glucose, diabetic and high blood sugar.
- Step 3: Visualize and validate the statistical model and clustering results.

IV. RESULTS

A. Statistical Analysis

The pre-processed dataset was applied in to the Minitab tool in three steps as follows;

1) Creating the statistical model

2) Testing Attributes

3) Removing the unnecessary attributes from the model

4) Validating the result

By creating the statistical model, the most effective and less effective attributes can be identified using P-value. Table 2 shows the P- values of each attributes.

A small P-value (typically < 0.05) shows the strong evidence against the null hypothesis. It means they are mostly affected to the logistic model. A large P-value (>0.05) shows the weak evidence against the null hypothesis. It means they are not mostly affected to the logistic model. There are 11 significant attributes. Among them the most significant attributes which have the minimum P- values were identified. Following are the attributes.

- Sugar Consumption
- BMI value
- Disease of pancreas
- Alcohol Consumption
- Background (Family History)

Table 2. P- Values of the attributes

Source	P-Value
Regression	0
Age	0.023
Sex	0.812
Sugar Consumption	0
BMI value	0.002
Disease of Pancreas	0.001
Diastolic Blood Pressure	0.027
Cholesterol Availability	0.04
Background (Family History)	0.006
Alcohol Consumption	0.006
Smoking	0.032
History of Heart Disease/ Stroke	0.031
Infection/Illness/skin rashes	0.023
No of Children	0.344
Dark, thick skin around neck	0.938

As the result of applying the binary logistic regression, it built up a statistical model which consists of two mathematical equations. They generate a way to predict the possibility of being a diabetic patient.

Equation 1. Calculating Y'

Y' = -12.71 + 0.0474 Age + 1.951 Sugar Consumption + 0.1499 BMI value + 2.167 Disease of Pancreas + 0.0542 Diastolic Blood Pressure - 1.081 Cholesterol Availability + 1.395 Background (Family History) + 2.89 Alcohol Consumption + 0.983 Smoking + 1.659 History of Heart Disease/Stroke + 2.24 Infection/ Illness/Skin rashes

$$P(1) = \exp(Y')/(1 + \exp(Y'))$$

Equation 1 consists of most relevant attributes which are most affected to the statistical model. The attributes values should be replace to this equation and then the value of Y' should be calculated well. Then the calculated Y' value should be replace to the equation 2 and should calculate the value of P(1). Prediction of being a diabetic patient depends on this P(1) value.

If the P(1) value is equal or greater than 0.5, The result is close to 1 (one). It means the prediction risk is positive. In other words, the person has the risk of being a diabetic 00person.If the P(1) value is less than 0.5, The result is close to 0 (zero). It means the prediction risk is negative. In other words, the person has not the risk of being a diabetic person. Positive range is identified as the value which is equal or greater than 0.5 and the negative range is identified as the value which is less than 0.5.

Statistical model was validated using two steps. Step 1 was done using randomly selected 25 instances from the data set of 200 instances. Step 2 was done using new 50 instances which also gain from government hospital in Kiribathgoda, Sri Lanka. Table 3 shows the performance of validated the statistical result.

Table 3. Performance of validated the statistical result

Validating Steps	No of correct predicting	No of wrong predicting	Accuracy result (%)
Step 1	23	2	92
Step 2	42	8	88

B. Clustering Analysis

The identified most suitable attributes (Sugar Consumption, BMI value, Disease of pancreas, Alcohol Consumption, Background (Family History)) were applied in to the WEKA tool. K-means clustering algorithm was used to cluster these data into 4 clusters. Table 4 shows the number of cluster instances and percentage of each cluster.

Table 4. Number of cluster instances

Cluster Number	Clustered Instances	Percentage
Cluster 0	55	28%
Cluster 1	28	14%
Cluster 2	59	30%
Cluster 3	58	29%

Table 5 shows how the 200 data categorized into the four clusters. Here, the clusters were identified as follows.

- Cluster 3: fasting glucose is greater than 300 mg/dl.
- Cluster 0: fasting glucose is equal and greater than 126 mg/dl and lower than 300 mg/dl.
- Cluster 1: fasting glucose is equal and greater than 100 mg/dl and lower than 126 mg/dl (impaired level).
- Cluster 2: fasting glucose is greater than 70 mg/dl and lower than 100 mg/dl

Table 5. Final cluster centroids

Clusters	Precision (%)	Recall (%)	Accuracy (%)
Cluster 0	98.18	84.37	94.50
Cluster 1	83.33	71.42	94.00
Cluster 2	72.88	82.69	87.50
Cluster 3	86.20	83.33	91.00

Clustering result was validated using precision and recall criteria. Table 6 shows the precision, recall and the accuracy of each cluster. Experimental results show that our approach performs effectively in clustering the data set.

Table 6. Performance measures of clusters

Attribute	Full Data	Cluster 0 (55.0)	Cluster 1 (28.0)	Cluster 2 (59.0)	Cluster 3 (62.0)
BMI value	0.3234	0.3442	0.2725	0.2379	0.4152
Disease of Pancreas	0.1	0.1273	0	0	0.4828
Background (Family History)	0.43	0	1	0	1
Sugar consumption	0.565	1	0	0	1
Alcohol consumption	0.1	0	0	0	0.4483

Figure 1 shows the visualization of each cluster in Weka. Here, cluster 0 is represent in blue, cluster 1 is represent in red , cluster 2 is represent in light blue and cluster 3 is represent in ash color.

According to the clustering data discovery, there are several decisions regarding diabetic disease. When the BMI value is increasing the level of diabetic is also increasing. If someone has covered the main 5 attributes in diabetic, he/she can have fasting glucose level up to 300. And also if someone has not covered any of these main 5 attributes in diabetic, he/she is in the normal level in fasting glucose. But even though any person has not covered all the main attributes but he/she has the relatives who have diabetic, there is a possibility to be in the impaired level in fasting glucose.

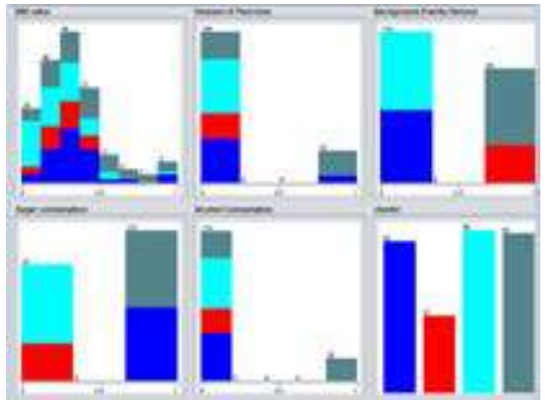


Figure 2. Visualizing the charts for most suitable attributes with fasting plasma glucose

Figure 2 shows how the most significant attributes (BMI value, Disease of Pancreas, Background (Family History), Sugar consumption and Alcohol Consumption) increased with the level of fasting plasma glucose.

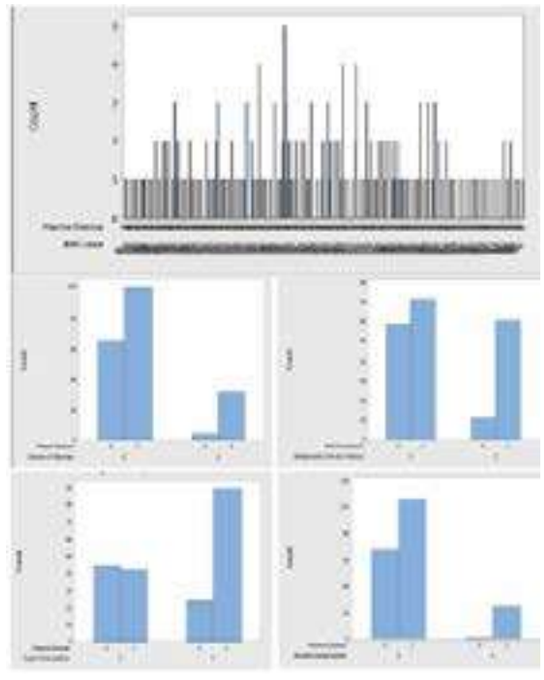


Figure 1. Visualization of Clusters

V. DISCUSSION AND CONCLUSION

The healthcare field is one of the most data enriched field in the world. As much as 30% of the entire world's stored data is generated in the health care environment. But unfortunately, the knowledge and learning getting from such kind of big data is very less. This research focused on those medical data. Every human being is once a patient who should deal with healthcare environment. So it is more important to holding a research actively with medical data. Aim is based on the value of predicting a disease and combining the computer system and information technology with the medical field. Predicting a disease is one of the major parts of dealing with medical data and the healthcare environment. Diabetic disease was selected here to predict in this research. Since diabetes is a chronic disease. An early prediction of the disease will save the patient life. Diabetes disease is the leading cause of death in the world over the past 10 years. In this research data mining was tremendously utilized to determine, analyse and further to visualize the useful information in diabetic. In this way data mining techniques are applied in medical data domain in order to predict diabetes and to find out efficient ways to treat them as well. Data mining is an interdisciplinary subfield of computer science is the computational process of discovering patterns in large data sets. It is also called as Knowledge Discovery in Databases which can help to support decision making in different fields including health care field. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The discovery of knowledge from medical datasets is important in order to make effective medical diagnosis.

In this research, statistics, clustering and visualizing methods were selected for analysing tasks. Fasting plasma glucose test reports of 200 people were used with all related 14 attributes were considered. Among them 11 attributes were identified as the most affected attributes. Statistical model was created to check the possibility of being a diabetic patient using binary logistic regression in Minitab. K – means clustering was used to categorized the dataset in selected attributes. Each cluster belongs to the level of diabetic.

In our future work, we plan to conduct a thorough evaluation of our proposed cluttering approach and statistical model by comparing with existing works.

ACKNOWLEDGEMENT

This study was done by the department of Computing and Information Systems in faculty of Applied Sciences in Sabaragamuwa University of Sri Lanka. The medical information was gained from Government Hospital in Kiribathgoda, Sri Lanka. Authors would like to thanks to the doctors of the Government Hospital in Kiribathgoda, Sri Lanka, especially Mrs Anoma Gamage for the valuable support and help.

REFERENCES

Ferreira, D., Oliveira, A. and Freitas, A., (2012). Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC medical informatics and decision making, 12(1), p.143.

Huang, Y., McCullagh, P., Black, N. and Harper, R., (2007). Feature selection and classification model construction on type 2 diabetic patients' data. Artificial intelligence in medicine, 41(3), pp.251-262.

http://apps.who.int. (2016). GLOBAL REPORT ON DIABETES. [ONLINE] Available at: http://apps.who.int/iris/bitstream/handle/10665 /204871/9789241565257_ eng.pdf;jsessionid =668AD6D16C59FD6CABA398693A61C579?sequence=1. [Accessed 10 July 2018].

https://www.niddk.nih.gov. (2013). Your guide to diabetes : type one and type two. [ONLINE] Available at: https://www.niddk.nih.gov/-/media/Files/Diabetes/YourGuide2Diabetes_508.pdf. [Accessed 10 July 2018].

Kelly, M. D. J. (2006). Diabetes. [Online] Available at: https://www.cdc.gov/media/presskits/aahd/diabetes.pdf. [Accessed 18 Jul. 2018].

Lowanichchai, S., Jabjone, S. and Puthasimma, T., 2006. Knowledge-based DSS for an Analysis Diabetes of Elder using Decision Tree. Faculty of Science and Technology Nakhon Ratchsima Raj abh at University, Nakhonratchasima, 30000.

Parthiban, G., Rajesh, A. and Srivatsa, S.K., (2011). Diagnosis of heart disease for diabetic patients using naive bayes method. International Journal of Computer Applications, 24(3), pp.7-11.

Rajesh, K. and Sangeetha, V., (2012). Application of data mining methods and techniques for diabetes diagnosis. International Journal of Engineering and Innovative Technology (IJEIT), 2(3).

Yuvarani, S. and Selvarani, R. (2016). An Analysis Of Decision Tree Models For Diabetes, International Research Journal of Engineering and Technology (IRJET), 3(11), 680 – 684.

pdf;jsessionid = 668AD6D16C59 FD6CA BA398693A61 C579?sequence =1. [Accessed 10 July 2018].

https://www.niddk.nih.gov. (2013). Your guide to diabetes : type one and type two. [ONLINE] Available at: https://www.niddk.nih.gov/-/media/Files/Diabetes/YourGuide2Diabetes_508.pdf. [Accessed 10 July 2018].

Kelly, M. D. J. (2006). Diabetes. [Online] Available at: https://www.cdc.gov/media/presskits/aahd/diabetes.pdf. [Accessed 18 Jul. 2018].

Lowanichchai, S., Jabjone, S. and Puthasimma, T., 2006. Knowledge-based DSS for an Analysis Diabetes of Elder using Decision Tree. Faculty of Science and Technology Nakhon Ratchsima Raj abh at University, Nakhonratchasima, 30000.

Parthiban, G., Rajesh, A. and Srivatsa, S.K., (2011). Diagnosis of heart disease for diabetic patients using naive bayes method. International Journal of Computer Applications, 24(3), pp.7-11.

Rajesh, K. and Sangeetha, V., (2012). Application of data mining methods and techniques for diabetes diagnosis. *International Journal of Engineering and Innovative Technology (IJEIT),* 2(3).

Yuvarani, S. and Selvarani, R. (2016). An Analysis Of Decision Tree Models For Diabetes, *International Research Journal of Engineering and Technology (IRJET),* 3(11), 680 – 684.

PARALLEL QUEUE OPTIMIZATION THROUGH COMPUTER AIDED SIMULATION AND QUEUEING THEORY - A CASE STUDY ON MATTA CANTEEN OF SABARAGAMUWA UNIVERSITY

K Banujan[1], S Anujan, KPN Rukshan, and MS Karunarathne
Department of Computing & Information Systems,
Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
[1]bhakuha@gmail.com

Abstract – A greater awareness and higher use of simulation technology caused many recent advances in the industry. ARENATM is a commonly used and convenient modeling and animating tool in the industry. It is grounded on object-oriented programming concepts and hierarchical modeling. Queuing is an obvious problem of the domestic canteens lacking the business philosophy of customer-centric in a market economy. When the simulation is over, we can find the model results and investigate the performance measures of our interest. As a case study, we selected the Matta Canteen of the Sabaragamuwa University of Sri Lanka. We simulated their process from 11.30 am to 1.30 pm. We applied the M/M/c queuing model for 2, 3, and 4 servers. The current system of 2 serving points has a 52.6 second waiting time in the queue, 3 serving points have 3.1 seconds and 4 serving points has 0.6 seconds. The average of customers waiting in the queue for the 2 serving point model is around 10 customers, but for the 3 serving point model, it reduced to 1 customer. According to our model, our recommendation is the 3 parallel servers to the current system. Our recommendation is further proved as the optimal model of the system using ARENATM and the queuing theory.

Keywords- Queue, Queueing Theory, Simulation, Food Serving, ARENATM

I. INTRODUCTION

Today, customers are much more conscious of cost perspectives before they choose a service or product. Performance metrics such as waiting time and activity costs are crucial for delivering a good-quality service at competitive prices. Generally, analytical tools are used to make decisions on staffing and pricing for a customer service process. But those tools have failed to consider the randomness and the system dynamics which are key parameters for queuing and variable costs analysis (Tumay, 1996).

Matta canteen is one of the main canteens which belong to Sabaragamuwa University of Sri Lanka. It is situated at the centre of the four faculties in the Sabaragamuwa University of Sri Lanka. Matta canteen caters breakfast, lunch and sometimes dinner for all the 7 days except some holidays. Further, they provide beverages like tea, plain tea, soft drinks etc. and short eats such as different bun varieties, mixture, vegetable rotti etc. in their open hours. They serve foods to students, academic staff, and non-academic staff. They provide dine in as well as take away services too. Approximately 2500 students, 300 academic staff members, and 250 non-academic staff members get a service from the Matta Canteen daily.

Queuing theory is concerned with the issue of waiting in the queue. Waiting is tedious and it should be avoided as it reduces the productive time of everyone within the system. Customers arrive to receive service by servers. Between arrival and start of service, there is a waiting time in the queue. Queues arise when the short-term demand for service exceeds the capacity of the server. It is most often caused by random variation in service times and the times between customer arrivals. Queuing and waiting time analysis is particularly important in service systems in order to improve their services which can attract more customers (Kelton et al., 2002).

Computer simulation tries to mimic real-scenario or hypothetical behavior on a computer in order to check how processes or systems can be improved. It also helps predict their performance under different circumstances and hypothetical scenarios (van der Aalst, 2010). Although simulation is typically considered as relevant and highly applicable to the real problem, but in reality, the use of simulation is limited. Most of the organizations have made an attempt to use simulation at some level in order to analyze their business processes. However, a very few are using simulation in a structured and effective manner (Van der Aalst et al., 2010).

By enabling enterprise-wide process simulation, ARENATM represents a most sophisticated advancement in simulation technology. It is a detailed system that takes into account all phases of a simulation project which starts with input data analysis to the output data analysis (Hammann and Markovitch, 1995). An impact on the various elements on the modeled system with different scenario combination could be analyzed using simulation. ARENATM assures a high degree of flexibility, various facilities for models of any level of complexity (Teilans et al., 2008). The modeling and executing system in ARENATM has the advantage of using the scaled formal layout of the real system.

II. LITERATURE REVIEW AND RELATED WORKS

ARENATM is a convenient modeling and animating tool that is based on concepts from object-oriented programming and hierarchical modeling. This kind of modeling tools is used by the second large community of system modelers. ARENATM is based on the SIMAN simulation language. It can be used for simulating different discrete and continuous systems in manufacturing, supply chain management, logistics, storing and other industries (Hammann and Markovitch, 1995).

The queuing theory is also widely using the historical random system theory, which examines the content of the system in the following three parts; (i) Behavior problems which mean that the queues probability is studied. It mainly deals with the queues, the length of distribution, and the waiting time distribution. Further, the busy period distribution is also looked into in both transient state and steady-state, (ii) Optimization problem which can be divided into the static optimization and the dynamic optimization. The static optimization problem is the optimal design, the dynamic optimization problem is the best operation of the queuing system and (iii) The statistical inference of queuing system is the third part that the robust model of a specified queuing system is determined for further investigating and analysing the queuing system (Xiao and Zhang, 2010).

In the study in "Solving waiting lines models in the airport using queuing theory model and linear programming the practice case" (Mehri et al., 2006), authors have conducted a case study on Tunisair Company at A.I.M.H.B during its' rush hour. They investigated the three parts of a queuing system (i) the inputs to the system, (ii) the queue or the waiting line, and (iii) the service readiness. After computing all the simulation results, the total expected costs are also studied by the authors. The total costs in their sense are the sum of the cost of providing service plus the cost of waiting time. By using these results they provided a feasible and profitable solution for the airline owner.

Ataeepour and Baafi (1999) showed in their research work, showed that ARENATM simulation models clearly show an enhancement of mine productivity with a dispatcher, especially when the truck fleet size is around the optimum value by using the Expected Delay Time concept. De Bruin et al. (2007) investigates the bottlenecks in the emergency care chain of cardiac in-patient flow. Their primary goal is to determine the optimal bed allocation over the care chain given a maximum number of refused admissions.

In "Modeling and simulation of emergency services with ARIS and Arena. Case study: the emergency department of Saint Joseph and Saint Luc Hospital" (Wang et al., 2009), authors have tried to solve the problem of overlong waiting time in emergency services which has the negative influence on healthcare quality. They resolved this by improving emergency services using modeling and discrete event simulation of the system process. They used IDS Scheer ARISTM and Rockwell ARENATM for their modeling purposes.

III. THE CURRENT PROCESS AND PROBLEMS FACED

Since Matta Canteen is located in the middle of the four faculties, most of the students, academic staffs, and non-academic staffs use this canteen for buying meals daily. More specifically for the students, Matta canteen is the key cafeteria for their meals. A considerable number of students usually come to Matta canteen on and on, especially in lunchtime (11.30 am to 1.30 pm). The lunch break usually falls between 12 - 1 pm and students have to walk a long distance to find another canteen to eat. This results in most of the students depending on Matta canteen for their having their lunch. Within the given short period, approximately 1 hour or 1 and a half hours, Matta canteen has to serve around 1300 students daily. Figure 1 shows the location of Matta Canteen.

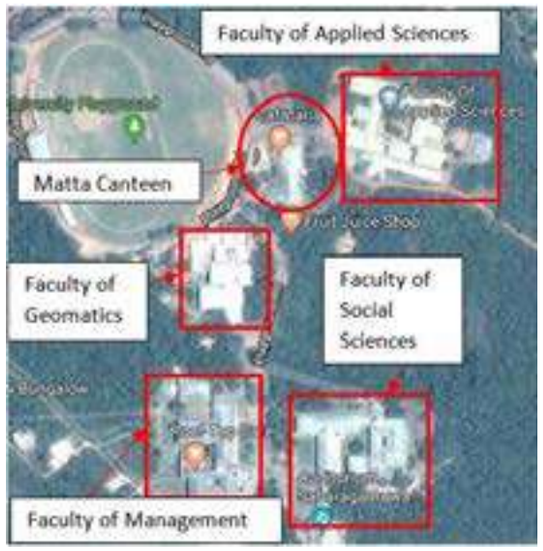


Figure 1. Map of Matta Canteen

Currently, the catering services are served by two service counters. After students get served from the service counters, they have to leave from the service counters and have their lunch by sitting in the seats available in the canteen. In our project, we are going to simulate the current situation and find alternative options, if they have 3 service counters and if they have 4 service counters for an effective service. Further, we are going to analyze the pros and cons of each three cases with the assistance of the computed results from the ARENATM simulation and by hand simulation. The Illustration of the current process is shown in Figure 2.

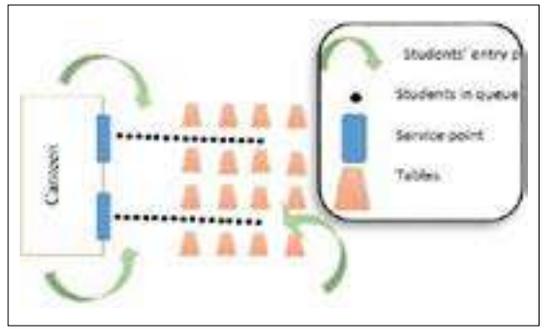


Figure 2. The current Process

For the time being, we took 11.30 am to 1.30 pm for our simulation because it is the most crucial time where students normally get rushed. Furthermore, there is a need for a proper solution in order to avoid the undesired consequences such as students miss the lunch due to long queues in a given short lunch break. Problems identified are; (i) There is a very long queue in the canteen very often, starting from the service counter to next end of the canteen, (ii) Students miss the lunch, and (iii) Students have to wait for a long time to get their meals. Therefore, because of the above-said problems, the aim of this work is to contribute to an improvement in the management of Matta canteen of the Sabaragamuwa University of Sri Lanka in order to reduce the waiting times for the students, academic and non-academic staffs.

IV. METHODOLOGY

As a first step for our simulation, we went to the Matta canteen and observed the process for a week (Monday to Friday). Currently, they are serving with two serving counters during the lunch hour. Based on our observations, we found that the queuing problem occurred during the time interval of 11.30am to 1.30pm. Therefore, we planned to simulate the canteen for that period. Then we analyzed those data gathered during the selected period to find the average inter-arrival time and average service time for a customer.

As a next step, we simulated the processes in the canteen by using the ARENATM Simulation Software. The advantage of using ARENATM was that the prior developed templates in various problem domains were readily available in it, and the reusability can be obtained. This means unlike conventional simulation systems in which all the modeling are hard-coded into the software by the vendor, ARENATM utilizes only soft-coded information which provides end users a freedom for reusable. We chose ARENATM because of these functionalities (Hammann and Markovitch, 1995).

Further, we applied queuing theory (Cooper, 1981, Gnedenko and Kovalenko, 1989) with M/M/c model with 2 servers, 3 servers, and 4 servers cases in order to verify with the ARENATM reports. Some notations used in the queueing theory are;

- λ : Mean rate of arrival.
- μ : Mean service rate.
- P = λ/μ: Utilization of the server.
- c : Number of servers.
- Pn : The probability of where n customers in the system.
- L : Average customers in the system.
- Lq : Average customers in the queue.
- W : Average time spent in the system.
- Wq : Average time spent in the queue.

When the mean service rate for each busy server is a constant for n all 1≥n, it is symbolized by μ. Here, when all c servers are busy, μn = sμ. Therefore, the expected inter-arrival time is 1/λ, the expected service time is 1/μ. Also, ρ = λ/(cμ) is the utilization rate for the service facility, i.e. the expected fraction of time as the individual servers are busy because λ/(cμ) represents the fraction of the system service capacity (cμ) that is being utilized on the average by arriving customers λ. The equations used in the M/M/c model is shown in table 1.

Table 1. M/M/c model equations

$\rho = \lambda/(c\mu)$
$P_0 = 1/\left[\sum_{i=1}^{c} c^i/i! + c^c/(c!(1-\rho))\right]$
$L = \lambda W$
$L_q = [(c\rho)^c/c!(1-\rho)^2] * P_0$
$W = (Wq + 1)/\mu$
$W_q = L_q/\lambda$

V. RESULTS AND DISCUSSIONS

Table 2 shows the results of the M/M/c queueing model for two, three and four servers for Matta canteen.

Table 2. M/M/c model outcomes with 2,3,4 servers

	C = 2	C = 3	C = 4
Average customers in the queue (L_q)	9.6449 customers	0.58 customers	0.1147 customers
Average customers in the system (L)	11.4783 customers	2.4134 customers	1.948 customers
Average time waiting in the queue (Wq)	0.8768 min (52.6087 sec)	0.0527 min (3.1638 sec)	0.0094 min (0.6257 sec)
Average time spent in the system (W)	1.0435 min (62.6087 sec)	0.2194 min (13.1638 sec)	0.1522 min (10.6257 sec)

Figure 4. 3D view of Matta Canteen three server counters

The current system of two serving counters has 52.6 seconds waiting time in the queue, three serving counters have 3.1 seconds and four serving counters has 0.6 seconds. By using this results we suggest that the current serving counters can be improved to three serving counters as the waiting time goes down drastically by 48 seconds.

Further, even though four serving counters reduce the waiting time up to around 2.5 seconds from three service counters to four service counters, it is not efficient to go for four serving counters because the cost to create a 4th serving counter cannot be overcome by reducing the waiting time.

Further, average customers in the queue also get drastically reduced. For the two serving counters model there are around 10 customers, but for three serving counters model it reduced to 1 customer. So three serving counters model will be the most efficient than two serving model.

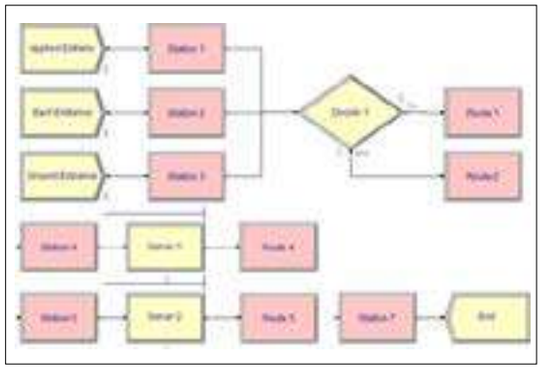


Figure 3. The layout of Matta Canteen for two serving counters

The assumptions used for analyzing the waiting lines are; (i) Customer arrivals into the system is Poisson distributed, (ii) Customer arrivals into the system come from an infinite population, (iii) Customer arrivals into the system are treated on a first in first out (FIFO) order and do not pause or revoke, (iv) Service times follow the negative exponential distribution or are constant, and (v) The average service rate is faster than the average arrival rate (Mehri et al., 2006).

The simulated results of the ARENATM show that the three serving counters are more effective than the two serving counters. Figure 3 shows the layout and the process of the Matta canteen for two serving counters in the ARENATM design view.

3 dimensional (3D) view of the simulated canteen at one particular time (at one instance) for three service counters is shown in figure 4 and two service counters is shown in figure 5.

Figure 5. 3D view of Matta Canteen 2 server counters

V. CONCLUSIONS AND RECOMMENDATIONS

Waiting lines and service systems are crucial indications of an effective service in the business world. In this study, the alternative ways of reducing the customers waiting time are studied using the queuing theory and simulation through ARENATM simulation software. The measure to reduce the time of customers' queues is obtained to achieve the goal of delivering good quality and effective service at the Matta canteen. According to the results we obtained through both queueing theory and ARENATM, the current method used by canteen i.e. two service counters can be improved to three service counters because the waiting time drops drastically when compared to two service counters. ARENATM delivers an integrated framework for building our simulation models conveniently and effectively.

Our future plans go as simulating for the Matta Canteen for a whole day of their working time i.e. 6.30 am to 6 pm. And also we are planning to do the cost estimation for our simulated results in order to strengthen our recommendations.

ACKNOWLEDGMENT

We would like to express our thanks to the owner of Matta canteen and the other staffs of Matta canteen for providing us with the great support of the study.

REFERENCES

Ataeepour, N. & Baafi, E. 1999. Arena Simulation Model For Truck-Shovel Operation In Despatching And Non-Despatching Modes. *International Journal Of Surface Mining, Reclamation, And Environment*, 13, 125-129.

Cooper, R. B. 1981. *Introduction To Queueing Theory*, North Holland.

De Bruin, A. M., Van Rossum, A., Visser, M. & Koole, G. 2007. Modeling The Emergency Cardiac In-Patient Flow: An Application Of Queuing Theory. *Health Care Management Science*, 10, 125-137.

Gnedenko, B. & Kovalenko, I. 1989. *Introduction To Queuing Theory. Mathematical Modeling*, Birkhaeuser Boston, Boston.

Hammann, J. E. & Markovitch, N. A. Introduction To Arena [Simulation Software]. Simulation Conference Proceedings, 1995. Winter, 1995. Ieee, 519-523.

Kelton, W. D., Sadowski, R. P. & Sadowski, D. A. 2002. *Simulation With Arena, Mcgraw*-Hill, Inc.

Mehri, H., Djemel, T. & Kammoun, H. 2006. Solving Waiting Lines Models In The Airport Using Queuing Theory Model And Linear Programming The Practice Case: Aimhb.

Teilans, A., Kleins, A., Merkuryev, Y. & Grinbergs, A. 2008. Design Of Uml Models And Their Simulation Using Arena. Wseas *Transactions On Computer Research*, 3, 67-73.

Tumay, K. Business Process Simulation. Proceedings Of The 28th Conference On Winter Simulation, 1996. Ieee Computer Society, 93-98.

Van Der Aalst, W. M. Business Process Simulation Revisited. Workshop On Enterprise And Organizational Modeling And Simulation, 2010. Springer, 1-14.

Van Der Aalst, W. M., Nakatumba, J., Rozinat, A. & Russell, N. 2010. Business Process Simulation. *Handbook On Business Process Management 1*. Springer.

Wang, T., Guinet, A., Belaidi, A. & Besombes, B. 2009. Modeling And Simulation Of Emergency Services With Aris And Arena. Case Study: The Emergency Department Of Saint Joseph And Saint Luc Hospital. *Production Planning And Control*, 20, 484-495.

Xiao, H. & Zhang, G. The Queuing Theory Application In Bank Service Optimization. Logistics Systems And Intelligent Management, 2010 International Conference On, 2010. Ieee, 1097-1100.

ARTIFICIAL NEURAL NETWORK BASED NEW CLASSIFICATION METHODOLOGY FOR IDENTIFYING KIDNEY DISEASE RISK LEVELS

KAA Chathurangi[1] and RMKT Rathnayaka[2]

[1]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

[2]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

[1]*ayeshachathurangi1110@gmail.com*

Abstract- The healthcare sector has vast amount of medical data which are still not properly analysed; especially, discovering useful information to predict future patterns is very limited. By using data mining techniques, the current study introduced a novel classification methodology and successfully applied it in Sri Lankan domain for Chronic Kidney Disease (CKD) classifications. The current study is carried under the two phases. In the first phase, Artificial Neural Network (ANN) method namely multilayer feed-forward neural network was used to detect whether a person has a risk of having a kidney disease or not and their risk level. In the second phase, a novel forecasting methodology is proposed using multiple algorithms, which is a combination of Random Forest algorithm and an ANN hybrid methodology to detect whether a patient has fallen into a CKD or not.

Keywords- Artificial Neural Networks, Data Mining, Random Forest

I. INTRODUCTION

The healthcare industry is producing huge amounts of data which need to be mined to discover hidden information for effective prediction, diagnosis, exploration and decision making. Analysing these huge amounts of data is complex and makes a huge challenge with available traditional methods.

As a result of these confusions, Healthcare Information Technology (HIT) has been developed as an interdisciplinary study of the design, development, adaptation, and application of Information Technology (IT) based innovations in health services for management and planning. HIT is a huge area comprising a multitude of components, solutions, and technologies.

The HIT plays a vital role in terms of improving the quality and effectiveness of healthcare, reducing healthcare costs and paperwork, improves the efficiency of both administrative and clinical processes, increases the accuracy of diagnoses, prevents medical errors, improve patient satisfaction and enabling better health outcomes.

As well as the benefits of HIT includes the ability to use data analytics and big data for effective management of population health plans and lower the occurrence of expensive chronic health conditions, the ability to share health data among academic researchers to introduce novel medical therapies and drugs, and the privilege of patients to acquire and use their own health data and work together in their own care with clinicians.

The HIT can be applied in several health domains like diabetes, heart disease, dengue, cancer and etc. Through these areas, huge amount of data are generated. The data mining techniques provide the methods and technology to generate useful information and patterns in large data sets for decision making and generate relationships amongst the attributes. Also it can be defined as the method of analysing data from various perspectives and summarizing it into information that are typically used to increase and enhance the revenue or reduce costs or to provide a new understanding and solution to a problems; especially, in several industries such as e-commerce, retail and social media.

In this case study, introduce a new classification methodology for kidney disease in Sri Lanka. As an initial step, the classification model predicts whether a person has a risk in kidney disease or not. As well as it shows risk level like high risky or low risky. Then the model predicts whether a person is fallen in CKD or not.

Kidneys in human body play an important role, with various functions that are critical to life. The main job of kidney is to filter and clean our blood. Kidney disease is an increasingly serious problem. When our kidneys are incapable to accomplish their functions properly; it may cause to occur kidney disease. Basically there are two types of kidney diseases that can be found in Sri Lanka.

A. Types of Kidney Disease

Chronic Kidney Disease (CKD) is kidney damage and a decline function that lasts more than three months. When someone is sick, injured kidney function, rapid changes in the cause, or taking certain medications, this is called Acute Kidney Disease (AKD). This can happen in ordinary kidneys or people who already have kidney problems.

B. Symptoms of Kidney Disease

Most common symptoms of kidney disease includes swelling of the body, specially noted in the face, feet, legs, hands and ankles, changes in urine output, frequency and colour, nausea, vomiting, itching of body, fatigue and exhaustion, sleeping problems, metallic taste in mouth, dizziness, breathing difficulty and chest pain, pain in the back and the sides and loss of appetite.

The rest of the paper is organized as follows. Section 2 explains about related works under this topic. Section 3 explains about the methodology of the research including technologies and techniques used in this study. Section 4 explains about the experimental design. Section 5 explains about experimental results. Section 6 is discussion and Section 7 ends up with conclusion.

II. RELATED WORKS

When considering about previous research publications it clearly shows this topic has already received a lot of attention in many researchers around the world. Different models and methodologies have been developed which is in related to this subject. In those models they have used several algorithms like Naïve Bayes, ANN, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbour (KNN), Random Forest and etc. Most of the research works are focused on finding a best algorithm or method that can be used in kidney disease classification and prediction and to identify the risk factors for kidney disease, symptoms of kidney disease, types of kidney disease and etc.

III. METHODOLOGY

The current study is carried out under two stages. In the first stage, The Back propagation algorithm which is a supervised learning method for multilayer feed-forward networks in the field of Artificial Neural Networks used to detect whether the person has a risk on having a kidney disease or not. Also it shows risk level like high risky or low risky.

This prediction was made by considering 11 attributes which are symptoms of kidney disease and that can be taken without any medical tests. The feed-forward neural networks are inspired by the information processing of one or more neural cells which is called as neuron. The fundamental of the Back propagation method is to create a given function by adjusting internal weightings of input signals to compose a desired output signal. The neural network model is trained using a supervised learning method. In here potential outputs of the algorithm are already recognized and the data set used to learn the algorithm is already identified with correct results (Arumawadu et.al, 2015; Arumawadu et.al, 2016).

In the second stage, a novel forecasting methodology is proposed using multiple algorithms which is a combination of Random Forest algorithm and an ANN hybrid methodology to detect whether a patient is fallen in CKD or not (Rathnayaka et.al, 2012). As an input data for this model, 30 attributes were used which is a collection

of general data about person, symptoms of kidney disease, results of medical tests and prediction results of Random Forest algorithm. Random Forest algorithm is a supervised classification algorithm (Rathnayaka et.al, 2014; Rathnayaka et.al, 2015).It is able to classify huge amount of data with an acceptable accuracy. At the training time it forms number of decision trees and outputting the class that is the mode of the classes output by individual trees.

A. Dataset Used

Data for this research was collected from special nephrology unit in provincial general hospital in Badulla. Dataset contains 108 instances and consists with 31 attributes. All these attributes represented in numeric or binary format.

B. Building Training and Testing Datasets

When modelling neural networks and other machine learning algorithms first it should be train. Then the trained models should evaluate. Therefore, dataset should divide into two as training and testing. Training data are used to optimize the weights in the neural network and other parameters in the model. Test data are used to evaluate the quality of estimates and forecasts respectively. Test dataset was not used for training models. Test data realistically simulated the model in the case where there was no information about the future. The test data is randomly selected. So that all data had an equal chance to participate in the selection process. In this study, dataset was separated as training dataset and testing dataset in 3 ways as shown in Table 1.

Table 1. Sample datasets for algorithm training and testing

Sample No	Training Dataset	Testing Dataset
1	60%	40%
2	70%	30%
3	80%	20%

C. Evaluate the Novel Model

In machine learning process, performance evaluation is an essential task. Some of the confusion matrix-based measures are used to evaluate the performance of the models constructed in this study using testing data. Such as accuracy, recall or sensitivity, precision, F1 score Receiver Operating Characteristic (ROC) analysis and Mean Absolute Error (MAE).

IV. PEXPERIMENTAL DESIGN

In the training and testing the models build in this study, machine learning approach was used. It is a methodology which uses to build mathematical models in order to understand data. Basically it can be divided into 2 phases namely build phase or modelling phase and operational phase. Training dataset is used in build phase. First features and labels should be extracted from the dataset. Then the selected machine learning algorithm should be trained until the model has learned enough. Once it learned, it should be saved. That saved finalized model can be used in the operational phase. Testing dataset is used in this phase. In operational phase you can ask from learned algorithm to explain newly observed data. This process is illustrated in the Figure 1.

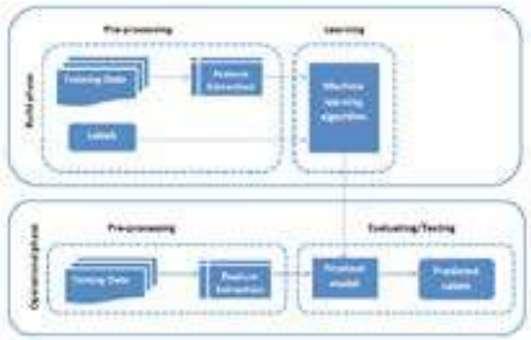


Figure 1. Machine learning workflow

In the process of ANN training; number of neurons in hidden layers and an epoch are adjusted until the target (known output) is reached. The training is stopped when the output result is consistence with the original result with least error rate. The output value of the models is in between the range 0.0 to 1.0. If the acquired output value is near to 1.0 then the person is having a risk on kidney disease or the acquired value is near to 0.0 then the person is normal person. The neural networks are trained and tested using three data samples and a neural network model with high performance was selected and saved. Then it can be used to perform the classification automatically for new pattern.

Machine learning algorithms are driven by parameters. Outcome of learning process of algorithms are highly depend on these parameters. So in here parameter tuning was applied to discover the best value for each parameter to enhance the accuracy of the algorithm or model. By repeating this process with a number of well performing models; optimum model can be selected. When training the Random Forest algorithm adjust the parameter values until best accuracy comes for three data samples. Then the highest accuracy shown model was selected and saved for model the CKD prediction model.

V. RESULTS

In the first phase, after a successful training of an ANN using 80% training data and testing that model using 20% testing data gives best performance. It gives 0.80952 accuracy and 0.19047 error rates for testing data. Final neural network model that constructed for make predictions for new data consists 11 input neurons in input layer, 14 neurons in first hidden layer, 9 neurons in second hidden layer and one output neuron in output layer. An epoch is set to 200. This model can be used to detect kidney disease and risk level. Figure 2 shows ANN structure of the model in first stage.

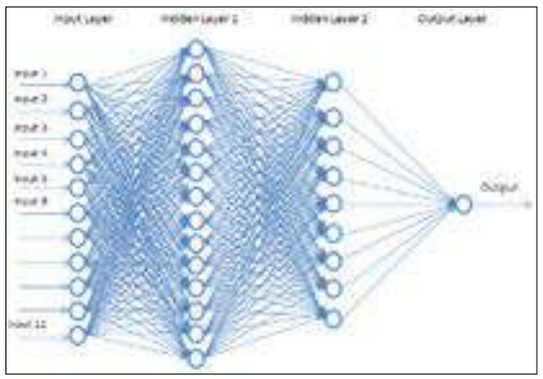


Figure 2. ANN structure for the model in first phase

In the second phase, a model with the combination of Random Forest and ANN including 30 input neurons in input layer, 10 neurons in first hidden layer, 9 neurons in second hidden layer, 6 neurons in third hidden layer and one output neuron in output layer was constructed for CKD prediction. It gives 0.81395 accuracy and 0.18604 error rate for testing data.

Figure 3 shows the performance comparison of model in phase 2 with other algorithms.

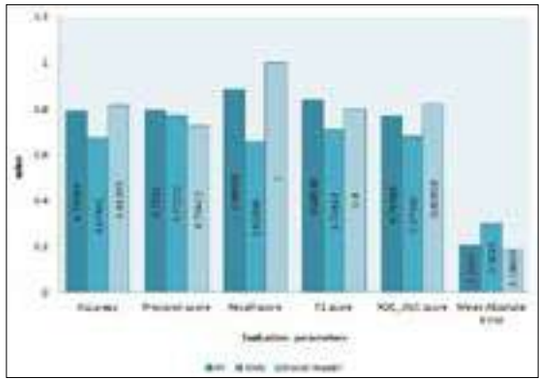


Figure 3. Performance comparison of algorithms

VI. DISCUSSION

This study introduces novel models for kidney disease prediction instead of finding a best or suitable algorithm that can be used to kidney disease classification and prediction. Figure 3 shows performance of CKD prediction of Random Forest algorithm, ANN and novel model built in the phase 2. The results clearly show the novel model built with the combination of Random Forest and ANN gives high performance in the CKD prediction instead of using Random Forest and ANN separately for CKD prediction.

VII. CONCLUSION

ANNs have been used in different medical fields and constitute useful techniques in clinical practice. This study introduces a novel model for kidney disease classification and prediction using Random Forest algorithm and ANN. Instead of using one method or one algorithm, the novel model was built by combining Random Forest and ANN. It gives high performance when comparing with algorithms separately. Artificial Neural Networks are frequently used as a strong discriminating classifier for tasks in medical diagnosis for early detection of diseases.

In the first stage, an ANN with one input layer, two hidden layers and one output layer was constructed to detect whether a person has a risk on having a kidney disease or not. In the second stage model was built by with the combination of RF and ANN with one input layer, three hidden layers and one output layer to predict whether a person is fallen in CKD or not CKD.

According to the health reports population of kidney patients in Sri Lanka is very high. But facilities required to treat them is very low. So this system will help to make the treatment process efficient. Sometimes persons want to know whether they are fallen into kidney disease or not, even though they have not fallen into kidney disease actually. So this system is very useful to them. Also this system is useful to kidney patients in initial stage. Also it reduces the cost and time.

ACKNOWLEDGEMENT

Success of this research depends on the encouragement and guide that I got from the many others. I would like to thankful Consultant Nephrologist and all the staff members of Nephrology unit in Provincial General Hospital, Badulla.

REFERENCES

Arumawadu, H.I., Rathnayaka, R.K.T. and Illangarathne, S.K., 2015. K-Means Clustering For Segment Web Search Results. *International Journal of Engineering Works*, 2(8), pp.79-83.

Arumawadu, H.I., Rathnayaka, R.K.T. and Illangarathne, S.K., 2015. Mining profitability of telecommunication customers using k-means clustering. *Journal of Data Analysis and Information Processing*, 3(3), pp.63-71.

Arumawadu, H.I., Tharanga, R.K. and Seneviratna, D.M.K.N., 2016. New Proposed Mobile Telecommunication Customer Call Center Roster Scheduling Under the Graph Coloring Approach. *International Journal of Computer Applications Technology and Research*, 5(4), pp.234-237.

Babu, A., Sumana, G. and Rajasekhar, M., 2013. Computer Aided Diagnosis of Polycystic Kidney Disease Using ANN. World Academy of Science, Engineering and Technology, *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering,* 7(12), pp.933-937.

Bala, S. and Kumar, K., 2014. A literature review on kidney disease prediction using data mining classification technique. *International Journal of Computer Science and Mobile Computing,* 3(7), pp.960-967.

Celik, E., Atalay, M. and Kondiloglu, A., 2016. The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning Methods. *International Journal of Intelligent Systems and Applications in Engineering,* 4(Special Issue-1), pp.27-31.

Gunatilake, S., Samaratunga, S. and Rubasinghe, R., 2015. Chronic Kidney Disease (CKD) in Sri Lanka-Current Research Evidence Justification: A Review. *Sabaragamuwa University Journal,* 13(2).

Jain, M.A.A.M.K., 2017. Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review. *International Journal Of Engineering And Computer Science*, 6(2).

Kalaiselvi, C. and Nasira, G.M., 2015. Prediction of heart diseases and cancer in diabetic patients using data mining techniques. *Indian Journal of Science and Technology*, 8(14).

Kaur, G. and Chhabra, A., 2014. Improved J48 classification algorithm for the prediction of diabetes. *International Journal of Computer Applications*, 98(22).

Khan, S.H., 2010. Predictive models for chronic renal disease using decision trees, naïve bayes and case-based methods.

Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V. and Fotiadis, D.I., 2015. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13, pp.8-17.

Kumar, K. and Abhishek, B., 2012. Artificial neural networks for diagnosis of kidney stones disease.

Patil, P.M., 2016. Review on Prediction of Chronic Kidney Disease using Data Mining Techniques. - *International Journal of Computer Science and Mobile Computing*, 5(5), pp.135-141.

Rathnayaka, R.K.T., Jianguo, W. and Seneviratna, D.N., 2014, October. Geometric Brownian Motion with Ito's lemma approach to evaluate market fluctuations: A case study on Colombo Stock Exchange. In Behavior, *Economic and Social Computing (BESC), 2014 International Conference* on (pp. 1-6). IEEE.

Rathnayaka, R.K.T., Seneviratna, D.M.K.N., Jianguo, W. and Arumawadu, H.I., 2015, October. A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models. In Behavioral, *Economic and Socio-cultural Computing (BESC), 2015 International Conference on* (pp. 54-60). IEEE.

Ratnayaka, R.K.T., Wang, Z.J., Anamalamudi, S. and Cheng, S., 2012. Enhanced greedy optimization algorithm with data warehousing for automated nurse scheduling system. *E-Health Telecommunication Systems and Networks*, 1(04), p.43.

Saumya, T.M.D., Rupasinghe, T. and Abeysinghe, P., 2014. A Literature Review in Data Mining Models Used for Survivability Prediction of Cancer Patients.

Shakil, K.A., Anis, S. and Alam, M., 2015. Dengue disease prediction using weka data mining tool. *arXiv preprint arXiv*:1502.05167.

A MACHINE LEARNING BASED SOLUTION FOR FINDING PERFECT MARITAL PARTNER

BKTP Wickramasinghe[1], DU Vidanagama, and N Wedasinghe
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]*wickramasinghe95@gmail.com*

Abstract- Marriage is a socially or ceremonially perceived joining between mates that sets up rights and commitments between those life partners. Finding a good marriage partner is one of the main reasons for the delay in marriage. Therefore, there is a need for a solution that can get user details and expected partner preferences and suggest proper matches based on their preferences. The objective of this paper is to discuss the necessity of the proposed model for Sri Lankans. The proposed solution will maintain user details and get user's preferences for their matrimonial partner. Based on the preferences, appropriate matches will be displayed to the user using a clustering algorithm along with the matching percentages. Horoscope of the user will be generated based on the planet details of the user. Furthermore, previous birth connection and 'dosha' identification will also be done. The proposed solution will also enable of sending messages to the matches and get email notifications about those matches. The final aim of this solution is to ease the matchmaking business by providing proper matches.

Keywords- Matrimonial Partner, Horoscope, Clustering, Matches

I. INTRODUCTION

Marriage is the union of two people. The definition of marriage varies around the world not only between cultures and between religions, but also throughout the history of any given culture and religion. A marriage ceremony is known as a wedding. In Sri Lanka, during the present century female age at marriage has increased almost by seven years. The delay in marriage has an enormous impact on the birth rate. Unlike in the west, where marriage is not necessarily the prosecutor of childbearing or the responsibility of running a household, in Sri Lanka procreation is almost entirely within the marriage. The United Nations World Fertility Report of 2003 reports that 89% of all people get married before the age of forty-nine. The percent of women and men who marry before the age of forty-nine drops to nearly 50% in some nations and reaches near 100% in other nations. (Un.org,2000) Finding a good marriage partner is one of the main reasons for the delay in marriage. In early times, people were not allowed to fall in love and get married. They had to find their matrimonial partner using traditional matchmaker approach ("Kapuwa") or Matrimonial advertisements. Even though those approaches were famous, there were some disadvantages. According to (Vreede-de Stuers,1969), the matrimonial advertisements that appear in English-language newspapers have attracted scholarly attention in recent years, and quite justifiably, for these items provide an abundance of information admirably adapted to statistical analysis of some of the variables determining mate selection. Yet the limitations of this material are also obvious. Verifying the accuracy of the contents is the main limitation. For instance, the beauty of a girl or the earnings of a boy can be exaggerated, and that may be misleading those who search the columns for an attractive candidate. Furthermore, the marriages happened based on these advertisements have led to divorce in most of the cases. With the advent of the Internet, a new channel in the form of matrimonial Web sites has emerged as an alternative way to find partners for marriageable members of the family. The introduction of technology in the form of matrimonial Web sites in an otherwise socially-enabled process provides the setting for a fascinating exploration of changing social mores and the interaction of technology and society. (Patnayakuni and Seth,2008)

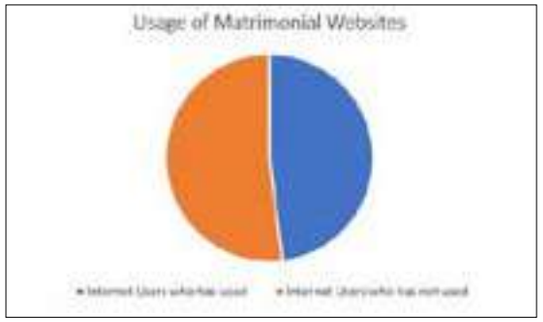


Figure 1. Usage of Matrimonial Websites

In 2015, India is estimated to possess 7% share of the total Internet users of the world and now is ranked 4th in the world in terms of Internet usage. 11% of the share is estimated to come from the Internet users who use matrimony sites (Figure 1). The total increase in growth of matrimonial sites is pegged to be 1500%. (Dasgupta, 2016) Interest in Online matrimonial reflects a new mindset among the urban youth and a major reason for that is the Internet and the large number of matrimonial sites offering a new landscape for romance and the search for a life partner without the traditional trappings or restrictions. Even though there are vast number of matrimonial websites, they do not support key features required in a Sri-Lankan marriage. By considering the facts from research papers and studying about existing systems, it could be understood that there is no proper system in Sri Lanka to find a matrimonial partner and fulfil the other requirements such as horoscope matching, wedding services etc. The main objectives of this research are: (1) Study the problems of existing matrimonial websites (2) Study the techniques and tools that can be used to develop the model (3) Present the developed model to solve the problem. The paper is organized as follows. Section 2 describes the background study. Section 3 describes the methodology. Section 4 presents the evaluation criteria and Section 5 presents conclusion about the framework together with directions in the future.

II. LITERATURE REVIEW

A. Existing Systems

There are vast number of matrimonial websites available in the market. Out of those websites huge number of websites belongs to Indian authors and dedicated to Indian users.

1) Shaadi.com: Shaadi.com is an online wedding service founded by Anupam Mittal in 1997. Shaadi.com began as Sagaai.com in 1997. Shaadi.com saw success over the next fifteen years as Internet adoption increased and people became more receptive to online matchmaking. (Shaadi.com, n.d.) Even though, shaadi.com is quite famous in India, it is not famous in Sri Lanka. It is basically created for India. If a Sri Lankan user is creating a profile, he cannot state the exact location. Location needs to be selected from the given set of values. Since it lacks Sri Lankan users, all the suggestions are for Indian candidates. It can be a huge limitation for those who strictly consider the religion, cast and nationality of their partner.

2) Jeevansathi.com: Jeevansathi.com is an Indian matrimonial portal owned by Info Edge. The website was bought in 2004 by Info Edge Ltd. The website has more than 6.1 million registered members as of March 31, 2014. The organization has around 230 employees working in 54 offices in 37 cities spread across the country, headquartered in Noida. It is listed in Bombay Stock Exchange.

Even-though it is quite famous, it does not provide you with perfect matches. Furthermore, Jeevansathi.com does not support Sri Lankan Candidates. Sri Lankan candidates cannot be searched using this website. (Figure 2) Horoscope matching service is also lacking in this website.

Figure 2. Browse Section of Jeevansathi.com

3) SimplyMarry.com: It is India's premier matrimonial service portal, promoted by the Time of India group. SimplyMarry.com offers one stop platform for online matchmaking that allows prospective Indian brides and grooms to meet and communicate regarding matrimony. Only, limitation in this site is lack of Sri Lankan candidates. Since it is an Indian site, it mainly focuses on Indian candidates. Furthermore, it lacks horoscope matching service. Out of these Indian websites, shaadi.com has been identified as the most user friendly, most used websites. The daily traffic rank trends (a measure of website's popularity) of these matrimony sites over a period of six months, are shown in Fig 3. Traffic rank is calculated using a combination of average daily visitors and page-views. The lesser the rank is of course better for a site in terms of popularity. Trend graph indicates daily traffic rank (charted over time), where shaadi.com leads currently. (Pal,2011)

Figure 3. Indian Matrimony Traffic Comparison

Even though shaadi.com is quite famous and trending, Sri Lankan candidates face a great deal in finding a perfect matrimonial partner through these Indian Websites. In the matrimonial market there are Sri Lankan matrimonial sites as well. They lack many important features.

4) SriLankanMatrimony.com: It is an exclusive matrimony portal catering to the special matrimonial needs of Sri Lankans across the globe. The portal offers several benefits to its members, primary being the pleasure of searching for a life partner within one's own community from across the globe at the click of a mouse. Even though there are some interesting features. There are limitations as well. To maintain a profile at SriLankanMatrimony.com, user needs to select a type of membership and needs to make the payment. Limitations in existing systems are summarized as follows. (Table 1)

Table 1. Limitations of Existing Systems

Existing Systems	Limitation
Shaadi.com	• Lack of Sri-Lankan candidates. • Incapability of stating Sri-Lankan towns as location,
Jeevansathi.com	• Lack of Sri-Lankan candidates. • Lack of Perfect match feature. • Lack of Horoscope Matching Feature.
SimpleMarry.com	• Lack of Sri-Lankan candidates
SriLankaMatrimony.com	• Lack of free services
LankaMatrimony.com	• Lack of Perfect match feature. • Lack of Horoscope Matching Feature.

B. Technology Used

1) Statistical Analysis Approach: The general approach of a statistical Matchmaker is to exploit that users are confirming preferences that they see fitting for themselves, without the need to call for an external expert or framework. A statistical Matchmaker tries to identify similarities between preference sets to come up with expectations or recommendations for a user, based on what user, identified as 'similar' by a distance function - expressed as their preference in the target context.

Dating agencies are using statistical methods to find good matches between people. They analyse the user profiles and their collected data about successful meetings to continuously improve their matching. (Segaran,2007)

In the statistical approach, appropriate algorithms from the field of machine learning and statistical analysis will be used for the matching of user profiles. A classical candidate approach is based on Support Vector Machines (Wassermann and Zimmermann,2011). Online dating services use statistics, data mining, and activity monitoring to provide appropriate matches; thus, differentiating their services and understanding the success of their product offering. (Smith,2005)

2) Clustering: Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics. (En.wikipedia.org,208)

(Gal, n.d.) has conducted a research on how k-means clustering allows to give the perfect match in dating apps. K-means algorithm was used to produce 15 clusters. A relatively high number of clusters was specified since the study deals with human behaviour, and more clusters are generally needed to see trends. Clustering algorithms find groups of observations that are similar to each other and different from the observations in other groups. For the study, both the k-means and EM clustering algorithms were run on the online dating data set to see if patterns could be found in the usage of and attitudes about online dating. In this case, the k-means algorithm was chosen as it had the most interesting results.

(Tausch et al., n.d.) has introduced a tree structure for efficient service matchmaking which was created by using the hierarchical clustering algorithm. Tree nodes represent a superset of all service descriptions in the leaves below. During query processing matchmaking can be restricted to the branches of the tree where tree nodes indicate overlapping between user requests and service descriptions. Good clustering of n service descriptions may improve retrieval time from O(n) to O (log n) for concise queries.

(Pei,2008) states that their experimental study reveals that the Probabilistic-based approach, employing the EM algorithm, yields better results than the Rule-based approach when the requirements to generate the probabilistic parameters are satisfied.

The results are summarized in Table 2. As per the literature review it was revealed that K-means clustering, and Statistical Analysis approach is suitable for the match-making process.

Table 2. Summary on Used Technologies

Machine Learning	Performance	Applicability
Statistical Analysis Approach	Good	Good
K-means Clustering	Good	Good
Hierarchical Clustering	Weak	Weak
EM Algorithm	Good	Weak
Canopy Clustering	Good	Weak
K-Nearest Neighbour	Good	Weak
Artificial Neural Network	Good	Weak

III. METHODOLOGY

A. Data Gathering

Qualitative and quantitative data required for designing the requirement specification for the new system were gathered by conducting a survey.

B. Data Analysis

The data which were gathered during the data collection phase described were analysed by using charts and presented in this section. Through analysing the data, was able to reveal the need of a proper matrimonial web solution for Sri-Lanka.

C. Approach

The users of this system are Candidates and Admin. There are five major different types of inputs from the candidates including Personal Details, Account Details, Education, Expected Partner Details and Horoscope Details. The outputs from the system are basically appropriate matches with horoscope matching details, previous birth connection details and dosha. Admin will generate user reports based on the candidate details.

D. Technology adopted

The input processing has been implemented using selected programming languages Html, CSS, JavaScript and Python. The technology that suits for the system development should be decided by considering the domain and the requirements for the system. It is important to identify most appropriate technological methodologies to satisfy the functional requirements and the non-functional requirements of the system in the system development procedure.

Proposed Matrimonial Solution is a web-based system. Time for the production, Efficiency and the performance of the system, Usability and Flexibility of the system and functionality of the system should be considered in developing the system. Programming language is the most important technical factor that should be focused in implementing the system. The programming language should in compatible with the development tools that are used in implementing the system. HTML, CSS, JavaScript were employed to develop the front end and Python was used to develop the backend. Flask, a python web framework was used and Pycharm IDE has been used.

E. Proposed Design

The architectural perspective of the developed system breaks into four main components as; software design, database design, module design and the interfaces. The gathered data from the analysis phase was used in creating the system's design.

1) Overall System Architecture

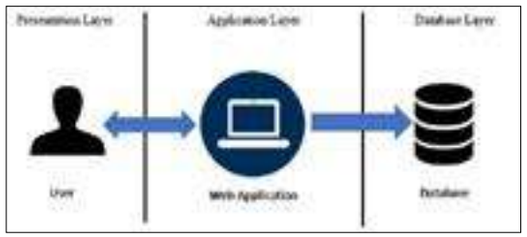


Figure 4. Overall System Architecture

Client layer: The user accesses the application through the client layer. The developed system provides access for the users in different user levels and each hold the interfaces which satisfy the requirements of each user type.

Application layer: Web Application belongs to the application layer. All the algorithms including porondam matching, clustering, dosha identification and previous birth connection identification are included in this layer.

Database layer: Database layer is responsible to manage the entire database of the system. The database will store the data enters to the system by the web application after the manipulation of the application layer.

2) Modular Architecture

The discussed modules in the software architecture will be elaborated in this section to show how the modules will be dealing with the users' actions in each layer. The main modules and their sub modules are given below.

i. **Login and Authentication Module**

Only the authenticated users can access the system. There are two user levels namely candidates and admin.

ii. **User Registration Module**

Candidates can register through the User Registration Module. Registration Module will be divided into five sub-forms namely Account Details, Personal Details, Education, Expected Partner Details and Horoscope Details.

iii. **Matching Process**

This module includes a clustering algorithm. Expected Partner preferences will be recorded during registration. Based on the preferences, appropriate matches will be displayed to the user using K-means Clustering technique. K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

iv. **Porondam Match Module**

Horoscope Details of a candidate will be stored in the database during registration. For the matches derived from the clustering algorithm, porondam will be matched. In this process five porondams are being matched.

- Nakath Porondam

  The most important aspect and crucial for a successful relationship between two persons. The Nakath porondam is the checking of whether the natures of the two marriage partners minds or in Sinhalese Sith are compatible.

- Rajju Porondam

  Physical bond of the marriage is through the certain parts of the body starting from the head down to toe. This porondam is about the compatibility of those parts which are producing sensations to the partners.

- Rashi Porondam

  Rashi means constellation. Rashi porondam is for checking and comparison of the status of the subconscious minds for compatibility.

- Nadi Porondam

  Nadi porondam considers the compatibility of the nerves of the two people, for an example if both partners get easily exited or angered then it is very difficult to maintain a long-term relationship.

- Graha Porondam

  This is about the influences of the destiny or in other words karmic factors. In other words, a person has certain control of his or her own destiny and others in the relationship and is checking to see whether this relationship is feasible to produce mutually beneficial outcomes.

\v. Previous Birth Connection Identification Module

To identify the previous birth connection, rashi should be generated separately, for both the horoscopes. Rashi of the horoscopes should be tallied. There are twelve rashis available. They are Aries, Taurus, Gemini, Cancer, Leo, Virgo, Libra, Scorpio, Sagittarius, Capricorn, Aquarius and Pisces.

vi. Dosha Identification Module

Apart from Porondam Match, dosha needs to be identified as well. Dosha will be identified individually for each of the candidate. If a certain horoscope contains a dosha, horoscope of the matching candidate should be able to cope up with the dosha.

vii. Message Sending Module

For the derived matches user should be able to send messages and interact with the matches.

viii. Feedback & Success Stories Module

User can submit feedback. Once a candidate meets someone to marry through this solution they can update the Status and once it happens the couple details will be displayed as a success story. Recent feedback and success stories of the couples will be displayed to the user.

ix. Email Notification Module

A trigger will be fired at the end of the day and the matches derived from the clustering technique will be sent to the user as email notifications.

x. Reporting Module

Admin will generate reports based on the candidate data. A dashboard will be generated.

The developed Machine Learning Based Web Solution is comprised of two main interface categorizations (user-levels) namely interfaces of Candidates and Admin.

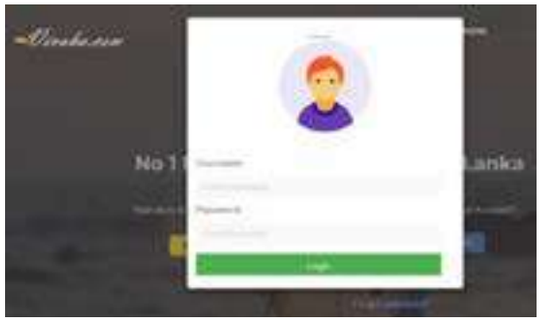


Figure 5. Interface of the Home Page
Figure 6. Login Interface

IV. EVALUATION

Summative evaluation was used as the evaluation method to find how the system functions and whether it is up to the expected level to fulfil the clients' requirements. At the system finalizing stage this evaluation is done to evaluate the product's stability. In summative evaluation a prototype with most stable build is shown to the client and the feedback is taken to find how far the system is success. In here the using prototype must be very much alike to the final product's functions and features.

The overall Evaluation of the product was carried to verify whether the system's final outcome meets the functional requirements of the users and the successfulness of the system tasks and the functions of each component are also evaluated here. This was done by comparing the functions of the new system against the problems and limitations addressed during the Requirements analysis process, also considering the functional requirements specified by the system specifications.

The prototype was given to ten people and they were asked to rate the system based on the following attributes their responses have been recorded and summarized as below (Table 3).

Table 3. Summary on Used Technologies

Parameter	Developed System
Efficiency	95%
Accuracy	75%
Availability	80%
Cost Reduction	98%
User Friendliness	97%

Furthermore, the prototype was given to five astrologers to check the accuracy of the horoscope algorithms. The results are summarized below (Table 4)

Table 4. Summary on Used Technologies

Algorithm	Accuracy
Porondam Match	64%
Previous Birth Connection Identification	96%
Dosha Identification	96%

According to the results, more than 95% of candidates have been satisfied by the mentioned parameters taken by the group for analysing the developed system and most of the astrologers state that the accuracy of horoscope algorithms are high.

V. CONCLUSION

In Sri Lanka, during the present century female age at marriage has increased almost by seven years. (Patnayakuni,2008) Finding a good matrimonial partner has become the major issue. Even-though there are traditional methods, they are not cost effective. Based on the results it is identified that the developed system performs well in finding a perfect marital partner. There are vast number of matrimonial websites available in the market. Many websites belong to Indian market and lacks Sri-Lankan users. The existing Sri-Lankan matrimonial websites are not up to the standard and lack many key features and do not support fee services. This paper proposes a web based matrimonial solution, which will generate appropriate matches based on the preferences along with the horoscope matching percentage. The authors will plan to further improve and refine the techniques, and to deal with the outstanding problems identified. The authors have also planned to generate an algorithm for fake detection of the feedback and the comments. These will be addressed as a further work.

ACKNOWLEDGEMENT

Authors would like to thank Mr. N Pathirage and staff of the Faculty of Computing at KDU, for their valuable and precious time, which is generously and highly admired.

REFERENCES

Un.org. (2000). *World Marriage*. [online] Available at: http://www.un.org/esa/population/publications/worldmarriage/worldmarriage.htm [Accessed 10 Feb 2018].

Vreede-de Stuers, C. (1969). The relevance of matrimonial advertisements for the study of mate selection in India. Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia, 125(1), pp.103-117.

Patnayakuni, R. and Seth, N. (2008). *Online Matrimonial Sites and the Transformation of Arranged Marriage in India*. [online] Available at: https://www.researchgate.net/publication/266908028 [Accessed 10 Feb 2018].

Dasgupta, S. (2016). "*A Suitable Bride: A Comparative Analysis of Matrimonial Advertisements of Indian Women Published in*

Newspapers and Matrimonial Websites". [online] Bing.com.

shaadi.com. (n.d.). [online] Available at: http://www.shaadi.com [Accessed 2 March 2018].

Pal, J. (2011). [online] Available at: https://www.researchgate.net/publication/256484632 [Accessed 2 March 2018].

Smith, A. (2005). Exploring online dating and customer relationship management. *Online Information Review*, 29(1), pp.18-33.

En.wikipedia.org. (n.d.). Cluster analysis. [online] Available at: https://en.wikipedia.org/wiki/Cluster_analysis [Accessed 15 April 2018].

Gal, M. (n.d.). *Online Dating and Marketing Using Machine Learning*. [ebook]

Tausch, B., d'Amato, C., Staab, S. and Fanizzi, N. (n.d.). Efficient *service matchmaking using tree-structured clustering*. [online] Academia.edu. Available at: http://www.academia.edu/2884228/Efficient_service_matchmaking_using_tree-structured_clustering [Accessed 18 April 2018].

Pei, S. (2008). *A Comparative Study of Record Matching Algorithms*. Edinburgh.

Segaran, T. (2007) *Programming Collective Intelligence*, O'Reilly, pp. 197-224 [10]

Wassermann,B and Zimmermann,G. (2011) "*User Profile Matching: A Statistical Approach*", in The Fourth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, p. 4.

CLOUD BASED POWER CONSUMPTION ESTIMATION FOR ELECTRIC VEHICLES

P.G.H.L.R. Fernando[1] and P.P.N.V. Kumara
Department of Computer Science, General Sir John Kotelawala Defence University, Sri Lanka
[1]*ruk.fernando93@gmail.com*

Abstract- Inaccurate range estimation is a major problem which comes with Electric Vehicles. Because of this many people face issues when planning long trips and short trips with limited battery capacity. To overcome this issue, it is necessary to have a better power consumption prediction algorithm which uses vehicle data and other dynamic environmental conditions. This paper is based on cloud based power consumption estimation system which uses linear regression in machine learning to obtain a better estimation based on above mentioned areas.

Keywords- Electric Vehicle, Range, Power Consumption, Estimation

I. INTRODUCTION

Electric Vehicles are now a trending vehicle type in Sri Lanka because of its environmental friendly approach, cost effective fuel consumption, comfortable interior design and cutting edge technology. According to the PUSL report related to electric vehicles (Public Utilities Commission of Sri Lanka, n.d.), from 2012 to 2017 there have happened more than 4000 EV registrations. Considering the size of Sri Lanka 4000 is a great number. But most of the people interested in EVs are not willing to buy an EV and the second hand market of EVs is rapidly decreasing. Industry research has uncovered that this is caused due to "anxiety felt by many drivers about the remaining driving range their vehicle can run before the next charge". This anxiety is mainly because of the current range and power consumption estimation algorithms do not accurately estimate the remaining driving range and required power to complete a journey. Range estimation technologies use limited data to calculate the estimated range such as battery health, state of charge, acceleration information, fixed auxiliary device power consumption and aggregated trip data. And current technologies do not use external environmental data such as traffic condition, speed limits, altitude and weather information to predict range. To overcome this anxiety, it is necessary to have an accurate power consumption prediction mechanism integrated into EVs.

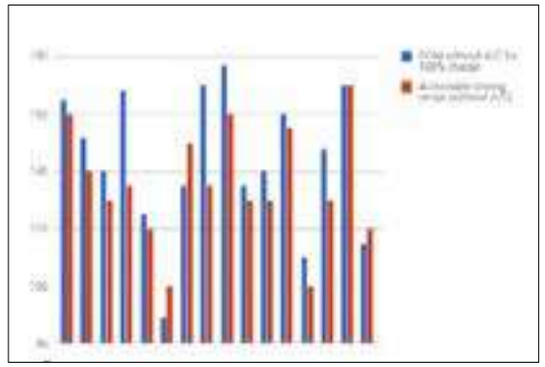


Figure 1. Difference between Estimated Range vs Achievable Range (km)

According to the survey based on Nissan Leaf 24kWh owners in Sri Lanka, Figure 1 shows that there's a considerable deviation on predicted range by inbuilt default prediction model in the vehicle with the achievable range for a full charge.

In order to predict accurate required power consumption to achieve a journey in an electric vehicle, it's necessary to have a prediction model which analyses battery consumption against achieved driving distance including various dynamic facts.

In this paper we propose a power consumption estimation framework which uses machine learning to identify the relationship in between various dynamic facts collected through OBD interface and GPS.

The rest of this paper is organized as follows: section II reviews the existing power consumption models; section III describes the range estimation architecture; section IV describes the methodology; section V describes about converting power consumption to battery percentage; section VI describes about switching between estimation models; section VII explains about the prototype and results; and section VIII concludes the paper and explains the future work.

II. LITERATURE REVIEW

The rest of this section is organized as follows: Prediction Model I; reviews the range prediction model proposed in the paper "Big-Data Framework for Electric Vehicle Range" by Habiballah Rahimi-Eichi and Mo-Yuen Chow. Prediction Model II; reviews the hybrid range prediction model proposed in the paper "Energy Consumption Prediction for Electric Vehicles Based on Real-World Data" by Cedric De Cauwer, Joeri Van Mierlo and Thierry Coosemans. Prediction Model III; reviews the macro range prediction model proposed in the paper "Energy Consumption Prediction for Electric Vehicles Based on Real-World Data" by Cedric De Cauwer, Joeri Van Mierlo and Thierry Coosemans. Prediction Model IV; reviews the range prediction model proposed in the paper "Range prediction of electric vehicles" by Viktor Schreiber, Axel Wodtke and Klaus. Prediction Model V; reviews the range prediction model proposed in the paper "Remaining Driving Range Estimation of Electric Vehicle" by Yuhe Zhang, Wenjia Wang, Yuichi and Keisuke Shirai.

A. Prediction Model I

This prediction model (Habiballah Rahimi-Eichi, n.d.) is a big data framework based on analysing real time data, previous driving patterns (Tseng, et al., 2012) and energy consumption of the vehicle. This model contains 5 sub models in order to identify Route and Terrain Information, Weather Information, Driving Behaviour Information, Vehicle Information and Battery Information.

By incorporating the results gathered from each sub module with a set of equations, authors have tested range estimation on simulated environment which was named SimBattery.

B. Prediction Model II

This model (Cedric De Cauwer, 2015) is based on the underlying physical principles of vehicle dynamics and kinematics. The total required mechanical energy at the wheels as a function of the kinematic parameters describing vehicle movement been expressed as a vehicle dynamics equation. And the equation includes five terms, each describing a contribution to the energy consumption. These terms describe, the rolling resistance, potential energy, aerodynamic losses, kinetic energy, and energy for the acceleration of rotational parts. According to the given equation by authors, the aerodynamic losses and rolling resistances are pure energy losses. And the potential and acceleration (kinetic) energy can partly be recovered by regenerative braking.

Battery consumption of the auxiliary devices is also used to predict the range of EV, and as described on the paper the consumption of auxiliary devices is based on the ambient temperature because of heating and air conditioning systems.

This model uses aggregated trip data to estimate range, bbut it does not use weather and traffic information which impacts the battery consumption. But since the estimation based on collected real-world data, authors have mentioned that weather and traffic information is inherently included in the calculation. Still two obvious factors which influence energy consumption, acceleration and weight information are absent in this prediction algorithm because the mentioned information cannot be extracted from the vehicle directly.

C. Prediction Model III

This model (Cedric De Cauwer, 2015) is an extended version of model II which was developed by same authors. The main idea of this model is to provide accurate range based on micro trips. This model has the potential to be more accurate than Model II as it does not apply averaging and so more information resides in the values of parameters. However, making prediction with this model is impossible dueto lack of available data, and to make this model applicable to range prediction of new trips, an additional correlation has to be done between characteristic values of these kinematic and physical parameters and external factors. However unlike Prediction Model I, accuracy of this model is tested on a real environment using data collected from Nissan Leaf 2012 24 kWh using OBD Port.

D. Prediction Model IV

This range estimation model (Viktor Schreiber, 2014) is based on preliminary studies and experiments of the energy flow and statistic method Design-Of-Experiments (DOE). According to the paper description, authors have divided driving cycles into 3 categories, such as motor ways, city and suburban area in order to analyse vehicle energy consumption through CAN signals. As a general model this contains 2 main sections as Energy Storage Model and Energy Consumption Model. Energy storage model analyse SOC (Stage of Charge) behaviour of the vehicle and Energy consumption model analyses energy consumption based on consumer and powertrain.

In consumer model, it provides the sum of all aggregated currents. Because usually consumers are periodically switching on/off, and they are running it permanently or they are not very dynamic.

Powertrain model analyse aerodynamic force, hill climbing force, rolling resistance, acceleration force and rotary inertia of powertrain to produce total power consumption of vehicle.

Apart from that this model uses GPS data to forecast route and prioritize routes. To forecast routes, it is required to obtain the destination from user. Based on the forecasted route other predictions like speed profile prediction, altitude prediction and driving time are made in order to improve the accuracy of range prediction. Overall the above mentioned areas are used in powertrain model to analyse energy consumption.

E. Prediction Model V

This range estimation model (Yuhe Zhang, n.d.) is developed considering 9 factors such as vehicle current location, remaining battery energy, road network topology, road grade, road link travel speed, acceleration and deceleration, wind speed, status of on-board electric devices and driver's driving style. This range estimation method is classified into 2 sections, such as rough range estimation and precise range estimation. Since processing part of this estimation model is done on a cloud server, all of the aggregated vehicle data are stored inside the server. It requires some processing power and time to calculate precise range, in order to reduce this situation, the authors have introduced rough range estimation model which takes less time to produce estimated range. From 100% to predefined level of battery percentage, rough range estimation model is used since drivers less cared about range when battery percentage is at satisfied level. But when battery percentage is at low level, drivers have anxiety about the achievable range from remaining battery percentage. In that scenario precise battery estimation model is used to predict range for remaining battery level.

III. RANGE ESTIMATION ARCHITECTURE

Like all other range estimation systems, this architecture also required to have all the standard data such as historical data, real time vehicle data including GPS locations (J. G. Hayes, et al., 2011), traffic data, weather information, battery information in order to predict range. Therefore, first we develop a framework to collect data from various resources and simplify collected data in order to apply machine learning. The range estimation system consists of 2 major prediction models; prediction based on previous matching data and machine learning prediction which uses a model trained by Google tensor flow applying linear regression. Figure 2 shows the diagram of data collection framework with major data collection nodes including vehicle information, weather information, traffic and route information.

Figure 2. Data Collection Framework

A. Vehicle Information

Major task of the vehicle information node is to collect real-time and static vehicle data such as battery capacity, state of charge, state of health, power usage under

different driving patterns (H. He, et al., 2012), (Tseng, et al., 2012), average power consumption for each user, power consumption of auxiliary devices including Air Conditioner etc. In order to collect the necessary information, we use ELM 327 (elmelectronics.com, n.d.) device which uses OBD interface to communicate with vehicle and retrieve CAR CAN Bus and EV CAN data stream. To retrieve data from OBD interface its necessary pass AT commands to the vehicle. Table 1 shows AT commands sequence which require to obtain data from Nissan Leaf Electric Vehicle.

Table 1. AT Commands Sequence to fetch Data (Tseng, et al., n.d.)

Leaf Setup	Decription
ATZ	reset all
ATLO	line feed off
ATSP6	set protocol to mode 6 (ISO-15765-CAN)
ATH1	set headers on
ATSO	set space off
ATCAFO	set auto formatting off
ATSH797	set header to 797
ATFCSH797	set flow control and header to 797
ATFCSD300014	set flow control and date to 300014
ATFCSM1	set flow control to mode 1
ATSH79B	set header to 79B
ATFCSH79B	set flow control and header to 79B

Likewise, through OBD interface it's possible to collect various meaningful information such as SOC (State of Charge), SOH (State of Health), Speed, Odometer, Motor Power Usage etc. Table 2 shows all the table of all the available data which can fetch over OBD interface of an Electric Vehicle.

Table 2. Available data over OBD Interface

Available data over OBD Interface			
Date / Time	Gps	Elevation	Speed
CP Min	Avg	Cell Pack	
Voltage	Volt	Temperature	Odometer
SOC	Ahr	Pack Volts	Pack Amps
Gids	SOH	CP Max Volt	Torque
Ambient Temperature	Motor Power	Anx Power	AC Power

CP – Cell Pack , Avg – Average

B. Weather Information

Although it's possible to fetch outside temperature from Car CAN Bus data stream, weather information node is used to fetch the temperature and other necessary information along the given route till reach the destination for given distance splits. In order to fetch weather information, we use openweathermap (OpenWeather, 2018) which is an open web based API. Through openweathermap it is possible to fetch temperature, pressure, humidity, wind speed with direction as degree in respect to north.

C. Traffic and route information

Route information such as distance to stop point, average speed, elevation, speed limits are the most important information set which requires to predict the power consumption of the electric vehicle. Using Google Maps' Directions API, it is possible to get all the mentioned information as an encoded string which is named as polyline. Through an available polyline decoder, it is possible to get GPS coordinates of the route and elevation. Since the prediction process happens inside the server we use PHP polyline decoder which is an PHP library under MIT and GPLv2 Licenses.

IV. METHODOLOGY

Major output of the system is a predicted numerical value which represents the average power consumptions of the given route. For an example if there are two locations namely A & B with a given distance in km, System will output the average required power in kWh in order to travel the distance from A to B. To predict this value, we propose 2 estimation methods which are

1) Find Previous Matching Data
2) Machine Learning (ANN) Based Prediction

Since the proposed system is a community based cloud application all the process happens inside the server and the historical data of the users are stored inside server. Since we can conclude that the dataset strength will be strengthened daily and the accuracy of the prediction models will get increased. But in order to protect privacy of the users, system will not store users' data specifically, but some portion of data will be used to analyse each driver's driving patterns specifically.

1) Find Previous Matching Data

This prediction model is based on finding nearest power consumption data which matches the current environmental condition based on the routes' GPS points. In order to apply this model, it is necessary to have previous power consumption records of the given route or similar set of data which can mimic the road condition of selected route.

It is necessary to fetch environmental changes including temperature, elevation and traffic status of the route and average driving speed of the route for given distance splits (default value is set to 200m). So if the total distance is 1km, system will collect traffic and environmental condition for each 200m, and finally it will have total of 5 records. Once the system collects data it will search for best matching records which mimic the required situation and get the power consumption value.

Figure 3. Sample code to find total power consumption

Figure 3 shows a sample code written in PHP laravel framework which can identify estimated power consumption for a given route by finding nearest possible behaviour patterns. In here the power consumption of auxiliary devices such as Air Condition and Audio System is not included and they will be added into final prediction by multiplying average required travel time since the required power for those devices remain constant most of the time if outside temperature remains unchanged. As shown in the figure 3, each feature has a threshold value of +/- 2 to maximize the availability of previous data records which mimic the current situation. In here we assumed that the weight of electric vehicle as a constant value in order to test the model, But it is a required feature in production release. This simple power consumption model provides accurate results for routes which have less elevation and temperature changes and for city areas.

2) Machine Learning (ANN) based prediction

Machine learning based power consumption estimation is also similar to the above mentioned prediction model. But instead of finding related previous data from the data set; this prediction model uses trained Artificial neural network model to predict data. Like model 1, this model uses various data as features additionally including wind direction, battery temperature and State of Health. This prediction model uses linear regression to train the neural network. Since there are many features provided as inputs, it is required to have a large amount of data records to improve the accuracy of the model. Figure 4 shows the expected vs predicted values of the ANN model which has been trained using a limited set of data. Since the weight of the dataset is low there is noise in the prediction. But overall the prediction is at a satisfactory level.

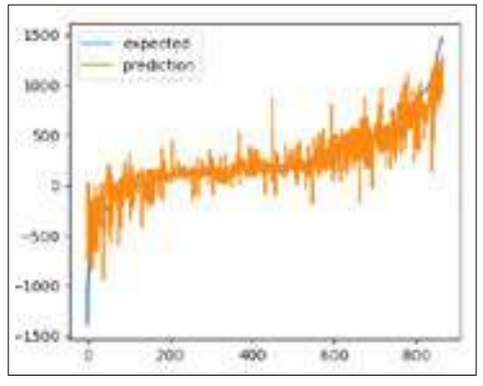


Figure 4. Expected vs Predicted Values of ANN Model

V. POWER CONSUMPTION TO BATTERY PERCENTAGE

SOH and SOC play a major role when converting prediction power consumption to battery percentage which is a human understandable value. SOC means the available battery capacity as a percentage value and SOH means the maximum capacity which can be held in battery as a percentage respect to the original battery capacity, or simply SOH represents the battery degradation of the electric car.

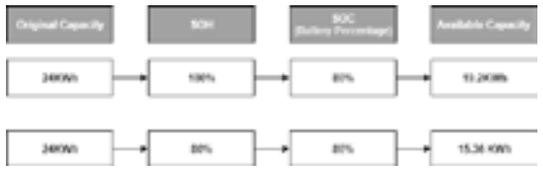


Figure 5. Find available capacity

By default, most of the electric cars cannot use advertised battery capacity. There is a threshold value. In a Nissan leaf AZEO 24kWh electric car, there is a 5% of threshold value, which means 95% of the capacity is usable (roughly 22.8 kWh).

Calculate remaining capacity of a Nissan Leaf 24kWh,

24*0.95*SOH*SOC

24*0.95*100%*100% = 22.8 kWh

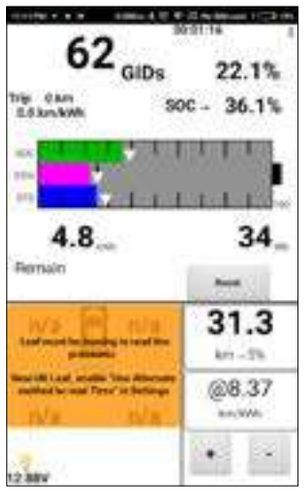


Figure 6. OBD Reading of a Nissan Leaf 24kWh

Figure 6 is an OBD reading of a Nissan Leaf 2014 24kWh taken from Leaf Spy mobile application which has 28% usable SOC (36.1% Total SOC) and 75.43% of SOH.

According to the above numbers the remaining capacity should be,

22.8*28%*75.43% = **4.81 kWh**
which is already displayed in the figure 6.

Calculating predicted required & remaining battery percentage,

R = Predicted required Power (kWh)
C = Current Battery Percentage (%)
B = Total usable capacity of a 100% SOH Battery (kWh)
SOH = state of health (%)

Required Battery Percentage = { R/(B*SOH) }*100%

Remaining Battery Percentage = C – { R/(B*SOH) }*100%

VI. SWITCHING BETWEEN MODELS

Since matching previous data is faster and accurate than ANN prediction, it has the highest priority over ANN prediction model. But due to low probability of available matching data in database there is a higher chance to switch prediction model to ANN based prediction model.

VII. PROTOTYPE & RESULTS

Figure 7. Power estimation GUI Front

Figure 7 shows the front end view of the power estimation application. Although there are several main inputs such as SOH, SOC, starting point, end point, power use of auxiliary devices etc., user just have to provide end location only. All other mention inputs will be read directly from the CAN Bus data stream over OBD2 interface. Since all collected data for the prototype test is based on a specific location, prediction accuracy for the specific location is higher than predicting required power for other location which has different weather and traffic conditions. But overall system provides accurate predictions although system has limited number of data records. Another important fact is that the estimated power consumption value is based on the traffic data provided by the Google Maps API and on average electric vehicle drivers' power usage values. In that case if a driver uses to drive aggressively or at higher speeds; this value might be not accurate.

VIII. CONCLUSION & FUTURE WORK

Cloud based power consumption estimation system for electric vehicles was introduced to overcome range anxiety of electric vehicle users by providing the value of required power consumption to achieve selected location or range by the user. This framework collects all the power consumption records from system users, traffic records, and weather changes. The system consists of 2 major prediction models and the model which uses Artificial Neural Network to predict power consumption is used to train its model on daily basis in order to improve the accuracy of the prediction. Prototype mobile application was developed to measure the accuracy of prediction, and as future works the prototype application will be developed for production and will collect more historical data in order to strengthen the dataset and the model to improve the accuracy.

ACKNOWLEDGMENT

We would like to thank all the people who participated in the survey on electric vehicle achievable range vs predicted range and Mr. Brandon Martin for publishing his mobile application as an open source project.

REFERENCES

A.Daziano, R., 2013. Conditional-logit Bayes estimators for consumer valuation of electric vehicle driving range. Resource and Energy Economics, 35(3), pp. 429-450.

Canonical Ltd, 2018. Ubuntu Server. [Online]
Available at: https://www.ubuntu.com/server

Cedric De Cauwer, J. V. M. T. C., 2015. Energy Consumption Prediction for Electric Vehicles Based on Real-World Data, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium;: Electrotechnical Engineering and Energy Technology, MOBI Research Group.

elmelectronics.com, n.d. ELM327. [Online]

Available at: https://www.elmelectronics.com/wp-content/uplads/2016/07/ELM327DS.pdf

Habiballah Rahimi-Eichi, M.-Y. C., n.d. Big-Data Framework forElectric Vehicle Range Estimation, NC, USA: Department of Electrical and Computer Engineering, North Carolina State University.

H. H., . C. S. & X. Z., 2012. A Method for Identification of Driving Patterns in Hybrid Electric Vehicles Based on a LVQ Neural Network. Energies, Volume 5, pp. 3363-3380.

H. Y., Tseng, F. & R. M., 2012. Driving pattern identification for EV range estimation. *Electric Vehicle Conference (IEVC)*, 2012 IEEE International, pp. 1-7.

J. G. H., R. P. R. d. O., S. V. & M. G. E., 2011. Simplified electric vehicle power train models and range estimation. *Vehicle Power and Propulsion Conference (VPPC)*, pp. 1-5.

Lanka, P. U. C. o. S., n.d. REGULATION OF THE ELECTRIC VEHICLE CHARGING STATIONS. [Online]

Available at: http://www.pucsl.gov.lk/english/wp-content/uploads/2017/09/Consultation-Paper-on-EV-Charging-Stations-and-Consumer-Rights.pdf

Martin, B., n.d. HappyLeaf. [Online]

Available at: https://github.com/blandman/HappyLeaf

Nissan, n.d. Nissan Leaf. [Online]

Available at: https://www.nissanusa.com/vehicles/electric-cars/leaf.html

OpenWeather, 2018. OpenWeatherMap. [Online]

Available at: https://openweathermap.org/

Tseng, C.-M.et al., n.d. Data Extraction from Electric Vehicles through OBD and, s.l.: s.n.

Viktor Schreiber, A. W. K., 2014. Range prediction of electric vehicles. s.l., Ilmenau University of Technology.

Yuhe Zhang, W. W. Y. K. ,. K. S., n.d. Remaining Driving Range Estimation of Electric Vehicle, s.l.: s.n.

WATER INTAKE RECOGNITION SYSTEM BASED ON PRESSURE SENSORS AND BLUETOOTH TECHNOLOGY

MAJ Wjesekera and WPJ Pemarathne[1]
Faculty of Computing, General Sir John Kotelawala Defense University, Ratmalana, Sri Lanka
[1]punsi111@gmail.com

Abstract- Dehydration is a very common problem especially among the elderly people and patients, and monitoring daily fluid intake of a person is vital to avoid dehydration and many other diseases. When we talk about the importance of water, it is an essential element of life. Automating the fluid intake monitoring can help to avoid the risk of losing the recommended daily fluid intake. An automated system can monitor and keep tracking the daily fluid intake and send reminders to the users and guide them towards wellbeing. In this research, a system is developed to help elderly people as well as patients to monitor their fluid intake. The system is developed with a special stand using pressure sensor and Bluetooth and an android application which provides records and reminders. It was observed that the system provides accurate results and it is a low cost solution.

Keywords- Fluid intake monitoring, Pressure sensor, Bluetooth

I. INTRODUCTION

Water is the fundamental element of life. It has proven that the approximately 60 present of body weight consists of water. Our body needs water in all its cells, organs and tissues to manage the body temperature and the functions. (Jequier E, Constant F, Eur J ClinNutr. 2010). Normally, body loses water when breathing sweating and digestion. So, it is vital to get rehydrated by drinking enough water and by eating food which contains much water. The water requirement depends on many factors like, climate, age, health problems and physical activeness. This has proved that drinking water is very important and that it is also important to take a recommended daily water intake. (D. Bunn, F. Jimoh, S. H. Wilsher, and L. Hooper 2014)

(Jaehyoung Yu Harnsoo Han 2008) measured the water level in a river or a tank such as applications related to flood and farming applications. The application that has been built to measure the water level is based on main 4 types such as pressure, heat, image and the 4 supersonic waves.

According to Welch, hydration is important for a good physical health of a person. Liquid intake from drinking water and beverages are major sources of hydration. We tend to forget to drink water because of our busy schedule and that may lead to unwanted problems. In this study, we will be developing a smart system to elderly people and patients to guide them with their daily water intake. The system consists of a hardware unit (smart measure) and an android application.

Pressure sensor is used in order to measure the liquid intake of patients. Then the readings where send to an android application, which will calculate the water consumption, mange records and give alerts. This system is simple, low cost and maintenance-free. Monitoring drinking behaviours of people living at home alone is important not only to ensure that they maintain an adequate fluid intake but also to identify the drinking patterns. This system can easily have adapted to patients to monitor their fluid intake.

In recent years a review of literature regarding the application of various wireless system techniques used in the area of measuring the water level as well as Hardware Design of Wireless Sensor Platforms which have been developed using such Water level monitoring techniques. Significant developments in the area of measuring liquid intake as well as the strengths and weaknesses of existing systems are also discussed the implementing of a remote measuring station in this the remote stations are considered as simple measuring units with a communication interface so that they may be operating under the control of base station. The advantages of this paper are that there are no mechanical parts required, remarkable accuracy and resolution, and disadvantage of this system is the water level monitoring is developed slowly and it required temperature companion (Daniels 2009). According to them they present a method to spot gestures when receiving data through sensors. The method is a natural way of getting continuous signals and is based on two stages. (H. Junker, O. Amft, P. Lukowicz, and G. Troster 2018)

A Wireless system for monitor and control of water level in greenhouse. They had used ZigBee network and several sensors nodes. The advantage is low cost and high network capacity. The ZigBee network for water irrigation control monitoring system. Here they used lots of sensors to monitor the water level of a tank and it was based from the signal that is coming from the sensors. (Morley JE, Miller DK, Zdodowski C, Guitierrez B, Perry HM 1998). According to (B.Y.Lee and B.Y.park, 2008) the pressure sensor is easy to use but it has some limitations where it should be replaced because of a breakdown which may occur by the high water pressure.

According to Zhou when we have to monitor and control the conditions of the environment such as temperature and humidity the main techniques used is WSN. This technique reduces the time when we have to monitor the environmental conditions. Also, the network technology that will be used is called ZigBee. This can be used in mining industry. (Zhou Yiming 2017). Here they use RFID Tags as a wearable sensor device to measure the liquid intake. Also, they use sliding window-based techniques to measure the activities (R. L. Shinmoto Torres, D. C. Ranasinghe, and Q. Shi 2013). In this review we provide the need for water and its importance for humans. Also, there are diseases related to dehydration, and how to monitor the amount of water consumed. (Bar-David Y, Landau D, Bar-David Z, Pilpel D, Philip M. U 1998).

According to Gender when the data has been measured and collected by the sensor the computer then takes the data from the database and monitor to see if the data is accurate, and if it is ready for communication. For this purpose, the computer chip is monitored by a micro controller which can measure the data that is being stored in the random-access memory and the Read only memory and other data monitoring software.

II DESIGN AND IMPLEMENTATION

A. The High-level Architecture of The System

The system is consisting of two parts. The first we design a hardware measuring stand, which is embedded with a pressure sensor, Bluetooth module and Arduino Microcontroller. Then we have developed and android application. The application takes the data from the hardware unit, calculate the measurements and provide user interface. Application is providing information to the users regarding the daily water intake and send them alerts of reminds to take water during the day to meet their requirements. Application also keep the past record to generate reports.

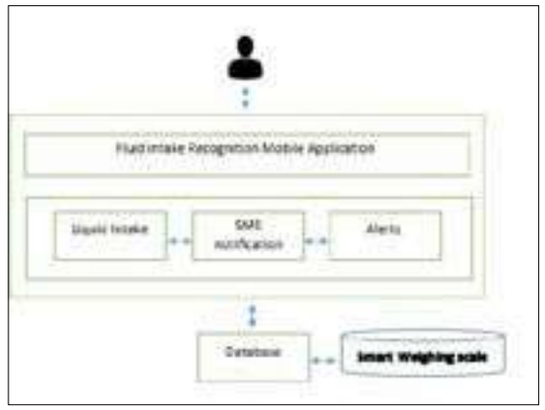


Figure 1. Overall System Architecture

Implementing android application will have the communicating media of a Bluetooth service. Only the registered users will be able to do the Measuring of liquid intake through implemented android application in the media of Weight Sensors.

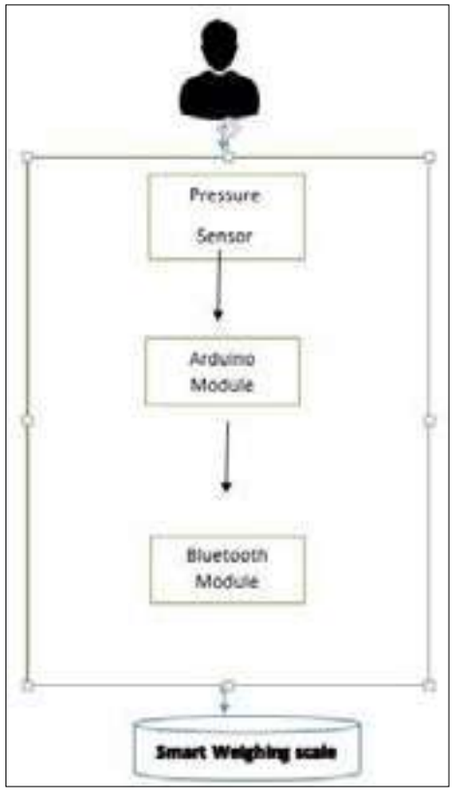


Figure 2. Identified Variables between Technologies

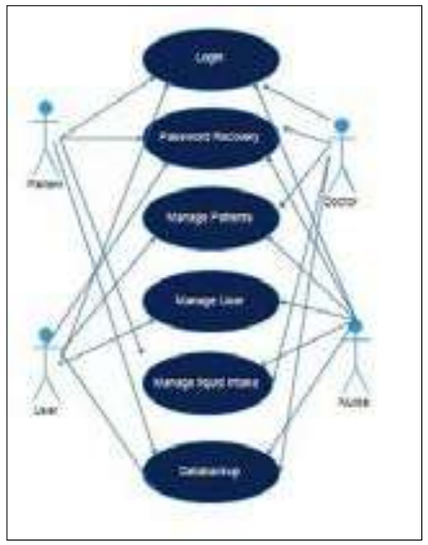


Figure 3. High Level Use Case

B. The design of the system core.

The core of the system is developed using Pressure Sensors, Arduino and Bluetooth Modules. However, the application should consist of user friendly and responsive interfaces which will allow user to access the platform on any mobile device. When we talk about the development of the mobile application, the interface module will consist of two basic Levels to input the data such as The Patient Level and the internal level.The interface will be for the patients to enter their details and check their status, and also the other interface will be for the authorized internal personal to enter the patient's records and see the status of the patients.

Figure 4. Sensor Input Data

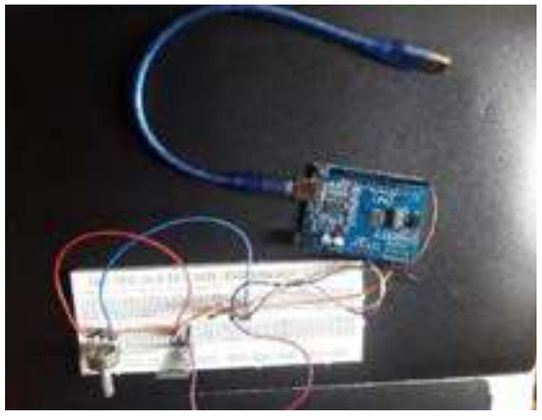


Figure 5. Hardware Interface

III. TECHNOLOGY USED

A. Use of Pressure Sensors

The reason to use the MS5837-30BA for the project is because The MS5837-30BA is because when we use a water proof pressure sensor the human interaction is limited and therefore we cannot get the accuracy of the data in a proper manner. When we talk about pressure sensors it introduces a good and accurate solution to measure the water intake of a person. Even though pressure sensor is easy to use, it also has certain limitations that has to be changed because of the pressure of the water.

B. Use of Bluetooth Module Connection

When we talk about wireless communication the first thing that comes to our mind is the Bluetooth connection. So, in order to communicate with a smart phone, and also in relation to the project we decided to use the Bluetooth module to transfer data from the pressure sensor to the mobile device. The reason to use this Bluetooth module is based on the following facts, such as Bluetooth consumes less energy than Wi-Fi and therefore it is easy to use.

Here we will be using the HC-06 is a class 2 slave Bluetooth module where it is used to transparent wireless serial communication. So, when it is connected to a Bluetooth device such as an android phone the operation will be transferred to the user, and all the data that is being received will go through a serial input and be transmitted over the air. It is easy to transfer data from multiple devices such as in this project, from the sensor you transfer all the date into one smart phone over a short distance.

C. Use of Arduino Microcontroller Technology

For the purpose of this project we used the Arduino – UNO because it can be easily used as both a hardware and software component. This consist of a circuit board and also it is built upon a readymade software called the Arduino IDE (Integrated Development Environment) which can be used to write the upload the code to the computer to the physical board. We need a microcontroller to measure the process from the pressure sensor and send it to the Bluetooth module, as well as to turn the Bluetooth module on and off when required. We use this platform because of its easy to use design development environment.

According to a study by KIipnisit is important to measure and monitor the physical conditions of a patient. When we monitor those conditions, we can get the internal data from the patient's body and get the accurate data of their health. And recognize the symptoms and help us to prevent them from dangerous body failure.

D. Android Studio

In order to make the mobile application we will be using the Android Studio Software in order to design interfaces.

Hardware interfaces are arranged as shown in the figure 5. The Programming Sensor Input are developed using Android Bluetooth Controller integrated with Bluetooth technologies which facilitates the need of responsive interfaces. Figure 4 shows the test data inputs taken by the hardware unit.

IV. HOW SYSTEM WORKS

The prototype model of the smart measure is shown in figure 5. After that we will connecting it to the Arduino using Breadboard and Jumper Wires. Next, we will take all the signals we get from the Arduino and connect it to the Bluetooth Module in order to have the communication with the android phone.

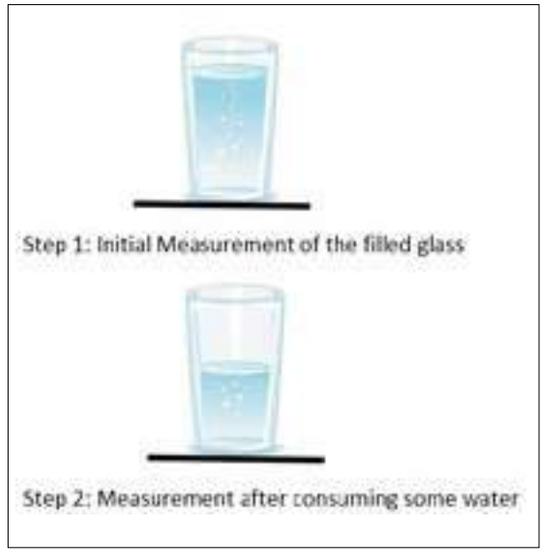


Figure 6. steps of measuring water consumption

Figure 6 shows the steps a person should follows when drinking water. First user has to fill the water and kept that on top of the measure as shown in step 1, then it will record the filled amount. After finish drinking user have to keep the glass on measure record data as in step 2. Then the system will calculate the water consumption using the difference (1 millilitre (ml) of water weighs 1 gram (g)) and keep records. This system automates the water intake, so it will be a great advantage rather than taking and keeping these records manually.

We will be recording all the data in the Database and sending all the relevant details to the mobile application so that every time the user drinks the water he/she will get the notification and the results of his behaviour. We will also be implementing a Mobile Application so that the patient can log on to the application and check the details whenever required. This system can be helped to elderly people who lives home alone, and patients need to monitor their water intakes.

- The users can log in to the system with authentication. Then the user can monitor the liquid Intake and browse the amount of water that they have consumed.
- This will also help the secondary users like gradients or medical officers to follow the records and analyze.

The system will provide the following outputs

- Daily water consumption
- Notifications regarding the remaining needs
- Water intake analysis report

V. CONCLUSION AND FURTHER WORK

When we talk about mobile applications and its development we see that it is one of the most technical future. When we talk about the future of businesses in order to win and live in the society they should always explore and come up with new ideas to improve their business and innovations. When we talk about hardware platforms it will reduce the work of the current users as well as new users and bring out a good solution.

Also, when we talk about good user experience we can use this and extend this to a wide area such as their mobile devices based on their location. Also, we can use this system in hospitals, where it would be easy for the doctors to check the amount of water consumed before being taken to the operation theater. In this study we use pressure sensors and Bluetooth modules to monitor the liquid intake of patients. Here we propose a smart weighing scale where we keep the cup to measure the amount of water consumed per day.

In conclusion we show the efficient and easy method of recognizing the amount of water intake consumed per day in order to have a good fluid balance life, based on pressure sensors and Arduino modules. We believe that this project will be very useful for patients, doctors as well as people who are interested in the medicine field. In future the system can improved to manage records on all types of fluid taken buy a person. This also can be modifying and improve to use in hospitals to monitor liquid intake and the output to create balance charts to help many patients.

REFERENCES

Abdullah Muhd Asran Bin (2008), water level in tank using sensor and PID controller, btech. Thesis, University Malaysia Saraswak, Saraswak,123p

Aufderheide S, Lax D, Goldberg SJ. Gender differences in dehydration-induced mitral valve prolapse. Am Heart J. 1995; 129:83–86.

Bar-David Y, Landau D, Bar-David Z, Pilpel D, Philip M. Urine osmolality among elementary schoolchildren living in a hot climate: implications for dehydration. Ambulatory Child Health. 1998; 4:393–397.

Y. Lee and B. Y. Park, "Development of high precision underground water level meter using a buoyant rod load cell technique," KSAFM, Vol. 1, 1999, pp. 1-5

D. Bunn, F. Jimoh, S. H. Wilsher, and L. Hooper, "Increasing fluid intake and reducing dehydration risk in older people living in long-term care: A systematic review," J. Am. Med. Dir. Assoc., 2014.

Daniels MC, Popkin BM. The impact of water intake on energy intake and weight status: a review. Nutr Rev. 2009

H. Junker, O. Amft, P. Lukowicz, and G. Tr¨oster, "Gesture spotting with body-worn inertial sensors to detect user activities," Pattern Recognit., vol. 41, no. 6, pp. 2010–2024, 2008.

Jequier E, Constant F. Water as an essential nutrient: the physiological basis of hydration. Eur J ClinNutr. 2010; 64:115–123.

Kim, J., He, J., Lyons, K., Starner, T.: The gesture watch: A wireless contact-free gesture-based wrist interface. In: ISWC. pp. 15–22 (2007)

Morley JE, Miller DK, Zdodowski C, Guitierrez B, Perry HM., III. Fluid intake, hydration and aging. In: Arnaud M, editor. Hydration throughout life: International conference Vittel (France) Montrouge: John Libbey Eurotext; 1998. p. 247.

N. ZachosA wireless network of remote measuring stations: application in water level monitoring" Conference: Industrial Electronics, 1995. ISIE '95., Proceedings of the IEEE International Symposium on, Volume: 1

Nicolaidis S. Physiology of thirst. In: Arnaud MJ, editor. Hydration Throughout Life. Montrouge: John LibbeyEurotext; 1998. p. 247.

O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing-based assessment of fluid intake," in Pervasive Computing and Communications Workshops. IEEE, 2010, pp. 298–303.

R. Bhattacharyya, C. Floerkemeier, and S. Sarma, "RFID tag antenna-based sensing: does your beverage glass need a refill?" in IEEE International Conference on RFID, 2010, pp. 14–16.

R. L. Shinmoto Torres, D. C. Ranasinghe, and Q. Shi, "Evaluation of wearable sensor tag data segmentation approaches for real time activity classification in elderly," in International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Japan, 2013.

Ramsay DJ. Homeostatic control of water balance. In: Arnaud MJ, editor. Hydration Throughout Life. Montrouge: John Libbey Eurotext; 1998. pp. 9–18.

Stookey JD, Constant F, Gardner C, Popkin BM. Drinking water is associated with weight loss. Obesity. 2008; 16:2481–2488.

Warner RR, Stone KJ, Boissy YL. Hydration disrupts human stratum corneum ultrastructure. J Invest Dermatol. 2003; 120:275–284.

Zhou Yiming Survey of Wireless Monitoring and Control System for Temperature, Humidity and Soil Moisture using ZigBee. International Journal of Electronics, Electrical and Computational System IJEECS ISSN 2348-117X Volume 6, Issue 8 August 2017

METHOD TO ENHANCE FEATURES OF BIOMETRICS SECURITY MANAGEMENT AND FINGERPRINT IDENTIFICATION USING LOW-QUALITY IMAGES

RSW Arachchi[1] and RL Dangalla
Department of Computing and Information Systems,
Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, Belihuloya.
[1]rumeshikasrimali@gmail.com

Abstract- Fingerprint identification becomes the most well-known biometric system in nowadays. The system uses special fingerprint points called miniature which is unique for every person. Many systems have been used for various algorithms to do their identification process. Low-quality fingerprints are an unavoidable problem which occurs due to various reasons such as deformations of the skin and dryness. This research proposes a set of algorithms and provides a suitable solution for problems occurred by low-quality fingerprint images. Miniature marking use triplets, segmentation using morphological operations were used as the novel changes in this system. The novel triplet miniature method solved the alignment problem, which is a significant factor in extracting miniatures. The standard methods of Gabor filter was used for the image filtering. The proposed system, coded using MATLAB was successfully implemented and working same as other recognition systems.

Keywords- Fingerprint identification, Lowquality fingerprints, Miniature, Alignment, Gabor filter

I. INTRODUCTION

Recognition of people using biometrics is an emerging trend in the modern world. Biometrics gained a massive attraction in the industry due to the need for security for a broader range of applications. Among these, biometrics fingerprint became more popular because it is the most practical biometric and also secure and comfortable to collect.

Fingerprint structures are divided into three major classes or patterns, namely Arch, Loop, and whorl. These classes are further divided into five subclasses called Plain Arch, tented Arch, Left Loop, Right Loop, and Whorl. When we consider the three main classes following details can be found. [1]

Pre-processing, feature extraction, and post-processing are three main steps in a fingerprint system. Following figure 1.1 shows a general block diagram of a fingerprint identification system.

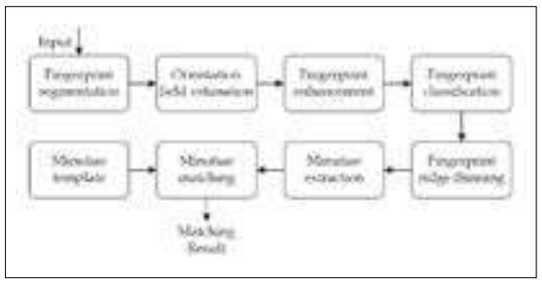


Figure 1.1 Block Diagram of a fingerprint machine Source : Ain Shams University, Egypt

The quality of the fingerprint images greatly affects the minutiae extraction process. To improve the performance of the system, many researchers put efforts into the image enhancement algorithms. Gabor filter and Fourier transformation are two of them. Gabor filter computes the response of eight oriented Gabor filters to determine whether a block belongs to the foreground or the background. Fourier Transform is a tool that decomposes the image into one and cosine components. It is a valuable image processing tool. When we input the image in the individual domain, the output will be in the frequency domain. Image compression, reconstruction, filtering & analysis are some applications using the Fourier transform.

The orientation field represents the orientation of the fingerprint. Estimate the orientation field the image was divided into 16*16 pixel blocks, and the gradient was calculated. From the gradient, information orientation angle is estimated. The orientation field is more important for latent fingerprints. Thinning process refers to the process of reducing the thickness of the lines as possible with minimum losses. This process is so important to identify the exact pattern of the fingerprint image.

One of the essential tasks considering an automatic fingerprint recognition system is the minutiae extraction from the captured image of the fingerprint. Due to imperfections of the acquired image, in some cases, certain minutiae can be missed by the extraction algorithm, and in other cases, spurious minutiae can be inserted. Therefore, it is important to choose an algorithm that solves all these questions.

A. Image matching techniques for low-quality images.

Matching fingerprints from a high-quality image is not a problem, but when it comes to low-quality images, it becomes harder. In the real world, not every fingerprint is clear. Therefore, it is important to consider how to deal with low-quality ones.

Cross-correlation is an efficient method for recognizing low-quality fingerprints. Correlation-based helps to work with images that are in low quality, damaged, incomplete or have shaped distortions. [31]

Phase -based image matching is another method for matching low-quality fingerprints. The use of phase components in 2D (two-dimensional) discrete Fourier transforms of fingerprint images makes possible to achieve highly robust fingerprint recognition for low-quality fingerprints. [26]

Phase only correlation can also be used for matching low quality fingerprints. This method makes possible estimate image displacement with 1/100 –pixel accuracy. [30]

There are many methods proposed for the extracting features from low-quality images that do not provide a perfect answer to the problem. The main aim of this research is to build a method that enhances extracting features from low-quality images.

II. METHODOLOGY

A. Fingerprint pre-processing

Fingerprint pre-processing contains four stages

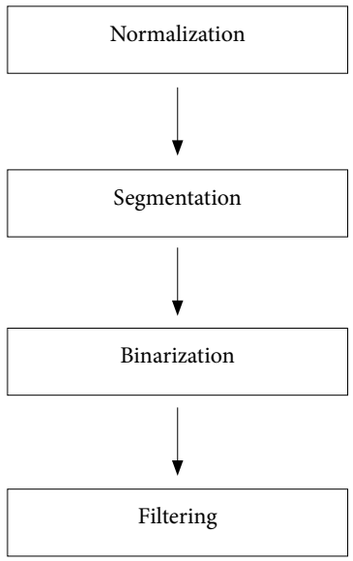


Figure 2. Pre - processing steps

1) Normalization: Normalization is done to identify the ridges and valleys easily. To do normalization. range between the ridges and valleys must be decreased. In the system, simple histogram equalization is done, which enhances the contrast of the image by transforming the values in the fingerprint image.
2) Segmentation: Orientation field represents the local orientation of the ridges contained in the fingerprint. From the normalized image orientation of the ridges are calculated in each block of the desired size. Three step approach fulfils the task.

i. Block direction estimation

ii. Segmentation by direction intensity

iii. Region of interest extraction.

Also, using morphological operations like erosion for extraction of ROI is introduced in this system.

3) Image binarization: Image binarization is a process which transforms the 8-bit grey image to a 1-bit image with 0 value for ridges and 1 value for furrows. To binarization, the image of a locally adaptive binarization method is performed. The image was divided into 16* 16-pixel blocks. Binarization of the image is done by an im2bw function within an inbuilt Matlab function.

4) Filtering: Gabor filters are used mainly in feature extraction, stereo disparity information and texture analysis in image processing and computer vision areas. Gabor filter is a special kind of bandpass filters. Bandpass filters are filters that allow necessary frequency bands and reject others. The sinusoidal-shaped waves of ridges and valleys very slowly in a constant local orientation. A bandpass filter, when tuned on a necessary frequency band, can remove unnecessary noise from the image. Therefore, it is better to use Gabor filters to filter these noises. Gabor filter is applied to an image if mainly concentrates on the edges and changes in the texture.[15]

Structure of the Gabor filter,

$$G(x,y,\theta,f) = \exp\left\{-\frac{1}{2}\left[\frac{x_\theta^2}{\sigma_x^2}+\frac{y_\theta^2}{\sigma_y^2}\right]\right\}\cos\left(2\pi f x_\theta\right),$$

$$x_\theta = x\cos\theta + y\sin\theta$$

$$y_\theta = -x\sin\theta + y\cos\theta$$

An im2bw function inbuilt Matlab function did binarization of the image. For the image preprocessing stage Histogram equalization and Fourier transform are used.

B. Miniature extraction

Miniature extraction contains following steps,

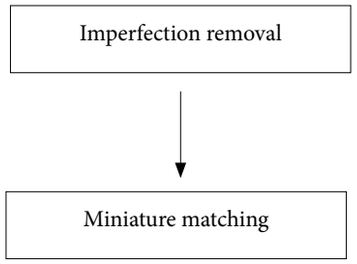


Figure 3. Post processing stages

1) Thinning: The second class of thinning algorithm is parallel. In parallel thinning algorithm results of the previous iteration is used for the decision for individual pixel deletion. This considers 3*3 neighbourhood around the chosen pixel. Rules of deletion applied according to the neighboring pixels. Entirely parallel algorithms have problems incorrectness. Therefore, they are broken into sub-iterations. In the method used midpoints of the black spaces are found and then joined to form a skeleton. This is fast but sometimes tends to produce noisy skeletons. After testing few thinning algorithms, the morphological thinning operation used because of high efficiency and good thinning quality.

2) Feature extraction: The system is based on matching miniature points, and it only considers ridges and bifurcations as a miniature. Extracting miniature is the most important part of the system. Due to various reasons, there will be false miniature. Various methods are introduced to remove or reduce false miniature.

After preprocessing the image extracting miniature, the image produces a large number of spurious miniatures. To remove those following heuristics like when there is a cluster of miniature in the same area are all removed. Accept the one in the centre and also if two miniatures are facing each other and there are no ridges between them both miniatures are removed from the set are used.

C. Post-processing stage

1) Imperfection removal: After thinning imperfections of the original image may remain to a certain extent depending on the image quality. This is spurious cause miniature. Therefore an algorithm must be made to remove lines not corresponding to ridges and connect broken ridges.

2) Miniature matching: Miniature matching is the most crucial step in the system. Here we match the miniature obtained from two sample fingerprints to check whether they are from the same person or not. Before doing the matching, we must consider the alignment of the image. Alignment of the image is important to match the fingerprint correctly because there can be plastic deformations in the finger. For this many systems use details of ridges or the Hough transform which are complicated to implement. There are three steps in the miniature matching stage,

i. Ridge correlation to specify reference miniature pair

ii. Align two fingerprint image

iii. Miniature match

In the system triplets of the miniature are formed, and distance from the other two miniatures and the angle between them are stored in a table. Then these tables are brute forced to do the matching. If we have n miniature n (n-1) (n-1) /6 triplets are formed for the fingerprint.

By this way, there is no need to rotate the image and do the matching. However, the problem is it takes time to do the matching. The system uses details from bifurcations because the probability of getting spurious details by the dry or low-quality image is a higher range. However, by using the good matching algorithm use ridge ending details also.

III. RESULTS

Statistical results were calculated with by matching images in the database. Genuine and imposter results show the scores obtained when we match fingerprints of the same person and when we match with the fingerprint of a different person. Imposter scores spread below 30%, while genuine score lies or spreads above that.

False Acceptance Rate (FAR) is the likelihood of the system to match or accept imposter fingerprints. In other words, FAR is the number of imposter scores that lies or spreads above the threshold value.

False Recognition Rate (FRR) is the likelihood of the system to reject genuine fingerprints. In other words, FRR is the number of genuine scores that lie below the threshold. 0.0329 equal error rate was found when the value of the threshold is 0.2623. Here, when the FAR and FRR is equal they are called Equal Error Rate. Accuracy of the system depends on EER. When EER is low, Accuracy of the system is high.

Following are some image results from the system,

Figure 4. Enhanced image

Figure 5. Binarization

Figure 6. Thinned

Figure 7. Feature extraction

The system successfully identifies whether two fingerprints belong to the same person or not. However, sometimes system produces erroneous results. Also, the computational time of the algorithm is high for a real-time application.

To make this method more efficient some places show scope for improvement. Image enhancement and miniature matching algorithm are some areas that need to be improved. The system produces perfect results when the image quality is high. But it's the real world, it may real differently because the fingerprint may be too wet or too dry. Also, deformations in the skin may occur in the real world.

IV. CONCLUSION

The main objective of this research was to build a security system that answers the low quality image problem. After a broad investigation, this system was build combining the best methods in the current world. This system also provides a solution for alignment of the image. The system follows all the necessary steps in an average identification system.

The existing system was upgraded using a new segmentation method, Gabor filter and also a new type of miniature and also with a database. These methods were used to increase the efficiency of the system for low-quality images. Also, some novel changes like using triplets of miniature to do the matching, segmentation using morphological operations are found in the system which was not reported in the literature referred.

Coding the system with MATLAB and going through all the stages of the fingerprint was helpful to understand key issues of fingerprint recognition. The algorithm used is not very robust and also vulnerable to plastic deformations and scaling. Various new algorithms have been found that gives better results.

The major challenge in the fingerprint recognition lies in the pre-processing of low-quality images. Still, a perfect answer for this has been not found.

ACKNOWLEDGMENT

I am glad to submit this paper of my research which was developing a fingerprint system for low-quality images. After successfully finishing my research I have so many people that I am grateful because they, guided me through this new path.

First of all, I would like to express my heartfelt thanks to the Head of the Department of Computing and Information Systems Dr. B.T.G.S.Kumara, who guided us throughout the research period giving valuable adviceand also to all the staff members of the department, applied science faculty, the Sabaragamuwa University of Sri Lanka for their wishes & the help for the completion of this project.

Words are inadequate in offering my thanks to my supervisor Mr. R.L. Dangalla, Lecturer, Department of Computing and Information Systems, Faculty of Applied Sciences, the Sabaragamuwa University of Sri Lanka for his assistance, guidance, encouragement through the research period and the completion of this project.

Last but not the least, I would like to express my gratitude to my family and friends who were with me all the time giving me the strength to make this happen.

REFERENCES

Al-Ani, M. S., 2013. A Novel Thinning Algorithm for Fingerprint Recognition. *International Journal of Engineering Sciences.*

Amira Saleh, Ayman Bahaa , A. Wahdan, n.d. Fingerprint Recognition. In: s.l.:s.n., p. 25.

Anil K. Jain,Arun Ross,Salil Prabhakar, 2014. An introdution to biometric recognition. Volume 14.

Anil K. Jain, Jianjiang Feng, Karthik Nandakumar, 2010. Fingerptint Matching.

Anon., n.d. Biometrics. In: s.l.:Cambridge computer laboratory, p. 16.

Anon., n.d. *Homepages.inf.ed.ac.uk*. [Online] Available at: https://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm

Anon., n.d. *Techopedia.com*. [Online] Available at: https://www.techopedia.com/definition/26990/biometric-system

Arsalaan. F. Rashid, Mehreen Lateef,Balbir Kaur, O. P. Aggarwal, Sajad Hamid, Neeraj Gupta, 2013. Biometric Finger Print Identification Is It a Reliable Tool or Not?. *Indian Acad Forensic Med*., Volume 35.

D. K. Isenor, S. G. Zaky, 1986. Fingerprint Identification Using Graph Matching. *Pattern recognition*, Volume 19.

D. Simon-Zorita, J. Ortega-Garcia, S. Cruz-Llanas , J. Gonzalez-Rodriguez, n.d. Minutiae Extractiob Scheme for Fingerprint Recognition Systems. p. 4.

Dileep Kumar, Yeonseung Ryu, 2009. A Brief Introduction of Biometrics and Fingerprint Payment technology. *International Journal of Advanced Science and Technology*, Volume 4, p. 14.

EMRE, A., n.d. Biometric Security Technologies. p. 17.

Feng Liu , David Zhang , Linlin Shen, n.d. Study on novel Curvature Features for 3D fingerprint recognition. Neurocomputing, p. 10.

Feng, J., 2017. Fingerprint Recognition. p. 118.

Fernando Alonso-Fernandez, Julian Fierrez-Aguilar, Javier Ortega-Garcia, n.d. An Enhanced Gabor Filter-Based Segmentation Algorithm for Fingerprint Recognition Systems.

G. Aishwariya, S. Kokilapriya, S. Adhithya, Dr. A. Grace Selvarani, 2017. Fingerprint Recognition for Android Application Data Retrieval. Volume 3, p. 4.

Gurpreet Singh , Vinod Kumar, 2014. Review On Fingerprint Recognition: Minutiae Extraction and Matching Technique. *International Journal of Innovation and Scientific Research*, Volume 10, p. 7.

Haiping Lu, Xudong Jiang , Wei-Yun Yau, n.d. Effective and Efficient Fingerprint Image Postprocessing. p. 5.

Haitham Hasan , S. Abdul-Kareem, 2013. Fingerprint image enhancement and recognition algorithms:a survey.

Hongshi Yan , Jian Guo Liu, 2008. Robust Phase Correlation based Motion Estimation and Its Applications.

IPriyanka Rani, IIPinki Sharma, 2014. A Review Paper on Fingerprint Identification System. *International Journal of Advanced Research in Computer Science & Technology*, Volume 2, p. 58.

James Wayman, Anil Jain, Davide Maltoni and Dario Maio, n.d. An Introduction to Biometric Authentication Systems. In: s.l.:s.n., p. 21.

Javier Galbally, S´ebastien Marcel Member, Julian Fierrez, n.d. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition. p. 17.

Kai Cao , Anil K. Jain, n.d. Automated Latent Fingerprint Recognition. p. 13.

Kaschte, B., 2005. Biometric authentication systems today and in the future. p. 13.

Koichi Ito, Ayumi Morita, Takafumi Aoki, Tatsuo Higuchi, Hiroshi Nakajima, and Koji Kobayashi, n.d. A Fingerprint Recognition Algorithm Using Phase-Based Image Matching for Low-Quality Fingerprints. p. 4.

Mridula, Priyanka, 2014. A Review on Classification of Fingerprint Images. *IOSR Journal of Electronics and Communication Engineering*, Volume 9.

NSTC, n.d. Fingerprint Recognition. p. 10.

Ravi Subban , Dattatreya P. Mankame, 2013. A Study of Biometric Approach Using Fingerprint Recognition. Volume 1.

Takita, K., 2003. High accuracy sub pixel Image registration Based on Phase- only correlation. *Fundementals*, Volume E86 -A, p. 10.

V.Karthikeyan , V.J.Vijayalakshmi, 2013. An Efficient Method for Recognizing the Low Quality Fingerprint Verification by Means of Cross Correlation. *International Journal on Cybernetics & Informatics*, Volume 2.

KNOWLEDGE SHARING SYSTEM FOR DENTAL EXTRACTION IN ORDER TO ASSIST DENTAL DOCTORS AND ASSISTANTS

K Banujan[1] and S Vasanthapriyan
Department of Computing & Information Systems,
Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
[1]priyan@appsc.sab.ac.lk

Abstract- Even though tooth extraction is one of the common surgical procedures in the dental field, it needs an extensive knowledge and practical experiences when handling the dental extraction equipment such as dental extraction forceps. Otherwise, it will be more complex or even in a worse case as it may cause damages to patients' mouth area. Hence, it is very important to have a sound knowledge of the instruments to be used, especially on extraction forceps. So, the knowledge of extraction forceps should be disseminated properly. After identifying this need, as a first stage, we gathered the information regarding the dental extraction forceps from the experts in the field. Then we started developing ontology as a second stage. Protégé OWL Ontology Editor 5.1 was used for this purpose. Finally, the developed ontology was evaluated in two folds; by using inbuilt tools and by ontology experts as an iterative approach. We strongly believe that our novel approach on dental extraction forceps ontology can support the dental students, dentists as well as their assistants to improve the knowledge and helpful in learning practices. Our next step is to model the ontology for whole extraction process and to develop a knowledge management system portal on dental extraction forceps.

Keywords- dental forceps, ontology, knowledge sharing

I. INTRODUCTION

A tooth is one of the hardest parts in most of the vertebrates which is a calcified structure and is situated inside jaws (Wikipedia, 2017). A dental extraction is the removal of teeth from its socket in the jaws (Kolosovas-Machuca et al., 2016). There are many reasons for dental extractions, but it is mainly done if a tooth has been damaged by decay or broken (Anyanechi and Chukwuneke, 2012). There are two types of extractions performed in the dental field:

i. A simple extraction, which is performed on a tooth which is visible inside the mouth. In a simple extraction procedure, the dentist will hold the tooth with specialized pliers called "extraction forceps" and move them front and back in order to loosen the tooth from the jaw before getting rid off the tooth (J. F. Cclyer, 1986, Wikipedia, 2017).

ii. A surgical extraction, which is a more complex like other surgical procedure.

The forceps is an exaggerated version pair of pliers. In general, forceps which are designed for the extraction of anterior (front) teeth in the maxilla (upper), the blades and handles are in the same line while for the maxilla posterior (back) teeth the handles form a curve with the blades. In forceps used for the extraction of mandibular (lower) teeth, the blades and handles are at an angle of approximately ninety degrees between them (J. F. Cclyer, 1986).

Sometimes different terminologies are used to express the same concept. Due to the unstructured, incomplete, general nature and varied formats of the information, the knowledge is not reaching everybody (Walisadeera et al., 2015). Further, computers need to understand the meaning or semantics of the information clearly. Semantic web enables this understanding of computers (Choksi and Jinwala, 2015). Ontologies are the powerful mechanism for representing knowledge presented in the semantic web (Vasanthapriyan et al., 2017b). Therefore, ontology can be used to find a response to queries within a specified context in the domain of dental extraction (Walisadeera et al., 2013, Vasanthapriyan et al., 2017b).

The aim of this work is to contribute to an improvement in the management and usage of dental extraction forceps in hospitals by developing an ontology-driven solution which organizes and describes clearly the related knowledge.

II. LITERATURE REVIEW AND RELATED WORKS

One of the commonest surgical procedure in the dental surgery is tooth extraction (Anyanechi and Chukwuneke, 2012). Dentists need to be very careful when performing tooth extraction. Even though enough efforts are applied to perform tooth extractions some accidents may happen when proper instruments are not used (Balaji, 2013). Sometimes it may lead to slipping of the forceps from the tooth to affect the other tooth or jaw or even partial removal of the tooth (Heimann, 1977).

All the tooth in human is not in the same shape and size. So, each tooth needs to be extracted using different types of dental extraction forceps. Choosing the appropriate extraction forceps is the important part for the protection of jaws and another neighboring tooth which will be affected if the forceps slips away while extracting. If the specific extraction forceps are not used for a tooth, then there will be more complications (Tevfik OLurel, 2014) such as incomplete extraction in which a tooth root remains in the jaw, prolonged bleeding, swelling, bruising, nerve injury or even extraction of the wrong tooth (Wikipedia, 2017).

Further, if proper extraction forceps are used by the dentist, his or her extraction will be easy as each extraction forceps are made by using the knowledge of physics. For example, extraction forceps are made smaller in size for children in order to apply less force and larger in size for adults in order to apply more force. If we take another example, the tooth on deep end in the jaw (wisdom tooth) needs a different mechanism to extract than the tooth in the front part of the jaw (incisors) (J. F. Cclyer, 1986). So extraction forceps are made "L" shape for the deeper end and straight for central area. Therefore, usage of specific extraction forceps for the specific tooth is very important for patients' health as well as for the easiness of the dentist.

Modelling knowledge by using ontologies in the medical domain is an active research field (Garcia-Valverde et al., 2014). Even though health sector is being supported by number of biomedical ontologies such as GALEN, the Unified Medical Language Source, the Systemic Nomenclature of Medicine which focus on general scope of the biomedical domain (Kuziemsky and Lau, 2010), and the Gene Ontology (GO) which is one of the earliest and most frequently used vocabularies (Hu et al., 2016), there are a very few ontology on dental domain on the health sector. Ontology-based systems provide reusable terminology resources and they can be used to improve the management of complex systems for different context information which can be captured and validated (Garcia-Valverde et al., 2014).

In order to integrate the knowledge, it has to be seamless and unaffected by the technological issues related to knowledge representation. In most of the knowledge ontologies, the experience of domain experts is key to design the ontologies (Chen et al., 2017). Semantic representations help in these issues and enable interoperability. For a particular domain, unambiguous description of the objects and their relationship can be described by using domain ontology (Rao et al., 2014). According to Gruber (Gruber, 1995), "ontology provides a structured view of domain knowledge and act as a repository of concepts in the domain". This structured view is essential to facilitate knowledge sharing, knowledge aggregation, information retrieval, and question answering. Therefore, ontology can be used in the domain of dental to find the specific response to queries.

The use of ontologies in the health domain mainly focussed on the representation and re-organization of medical terminologies. The most significant benefit that ontologies in the health sector are its ability to support the integration of knowledge and data (Pisanelli, 2004). Even though ontologies are used in the information system (IS) design, the ontology development in the health sector is more challenging because of its complexity and the level of detail in it (Kuziemsky and Lau, 2010). Even though there is some previous work which has evidenced knowledge sharing methods for various domains such as software testing (Vasanthapriyan et al., 2017b, Vasanthapriyan et al., 2017a), economics (Yoo and No, 2014) etc., a very little

research into dental knowledge sharing using domain ontologies has been conducted.

Having discovered this research gap we have focused on our attention on developing a dental extraction forceps ontology to represent information needs according to tooth extraction context. That is, we intend to develop an ontology-based knowledge framework to manage extraction forceps-related knowledge. This would assist the doctors and their assistants in the dental hospitals to manage extraction forceps knowledge.

III. METHODOLOGY

Our main focus is on simple extraction, more specifically in extraction forceps because these extraction forceps plays an important role in tooth extraction (J. F. Cclyer, 1986). Grounded theory was used for data collection. Two dentists with extensive knowledge of dental (mainly on extraction) and an expert on ontology engineering took part. After interviewing them, the dental extraction problems were identified. The Competency Questions (CQs) were developed after collecting needed information. CQs are a set of questions that the ontology must be capable of answering using its axioms (Grüninger and Fox, 1995). CQs work as requirement's specification of the dental extraction forceps ontology. Our ontology aims to answer competency questions. Some of them shown in table 1.

We get the relevant data in order to answer these CQs through an extensive literature survey and expert collaboration. We categorized the "Person" into two; "Adult" and "Child". The "Parts" of the tooth is divided into "Crown" and "Root". "Positions" also categorized as "Upper", "Lower", "Left" and "Right". Tooth have "Specific Names". They were classified into "Molar", "Premolar", "Canine" and "Incisor". There are three international standard systems for naming teeth:

i. The universal numbering system,

ii. The palmer notation method and

iii. The two-digit FDI world dental federation notation.

In this paper, we followed two-digit FDI world dental federation notation developed by International Association for Dental Research. It provides a system for designating teeth or areas of the oral cavity using two digits (Park and Kim, 2006). We declared these notations into "ToothNotation" class. Since there are no standards for forceps classification, we formalized the forceps into two main categories; Crown Forces and Root Forceps. Here what we meant by the crown is the full tooth which includes both parts; crown and root. Further, we divided each forceps into many subclasses. The high-level class hierarchy is shown in Figure 1. The ontology was implemented by using the Protégé-OWL Ontology Editor 5.1. Part of the dental extraction forceps ontology is shown in figure 2.

Since we are designing with OWL 2 Web Ontology Language for the semantic web, we use Description Logic (DL) which is a decidable fragment of FOL for our scenario. The competency questions were evaluated to check whether the developed ontology meets the dentists' requirements in the design process. In order to query the ontology, the DL expressions have been used. For this purpose, we used the DL query facility which is available in Protégé OWL Ontology Editor 5.1. Some of the DL query and their answers are shown in Table 2.

Table 1. Font sizes for this publication

Competency questions
What is the position of central incisor of a child in the mouth?
Which tooth is in the upper left side of an adult?
Which extraction forceps are needed to extract the root of left second premolar of an adult?
Which extraction forceps are needed to extract the normal left central incisor of a child?
Which teeth are in the upper left side of an adult?
What is the tooth of lower left central incisor of a child in the mouth?
Which extraction forceps are S-shaped?
What type of movement is applied to the third molar of the upper left side of an adult
What type of dentition is for a child?

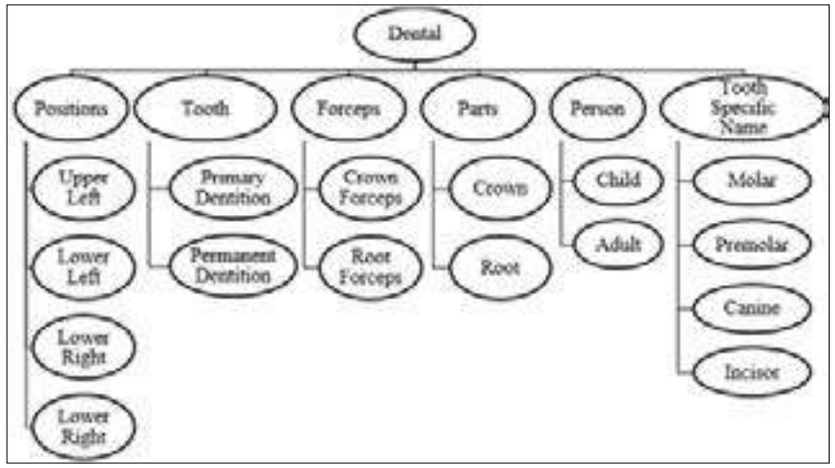


Figure 1. The high-level class hierarchy of dental extraction forceps ontology

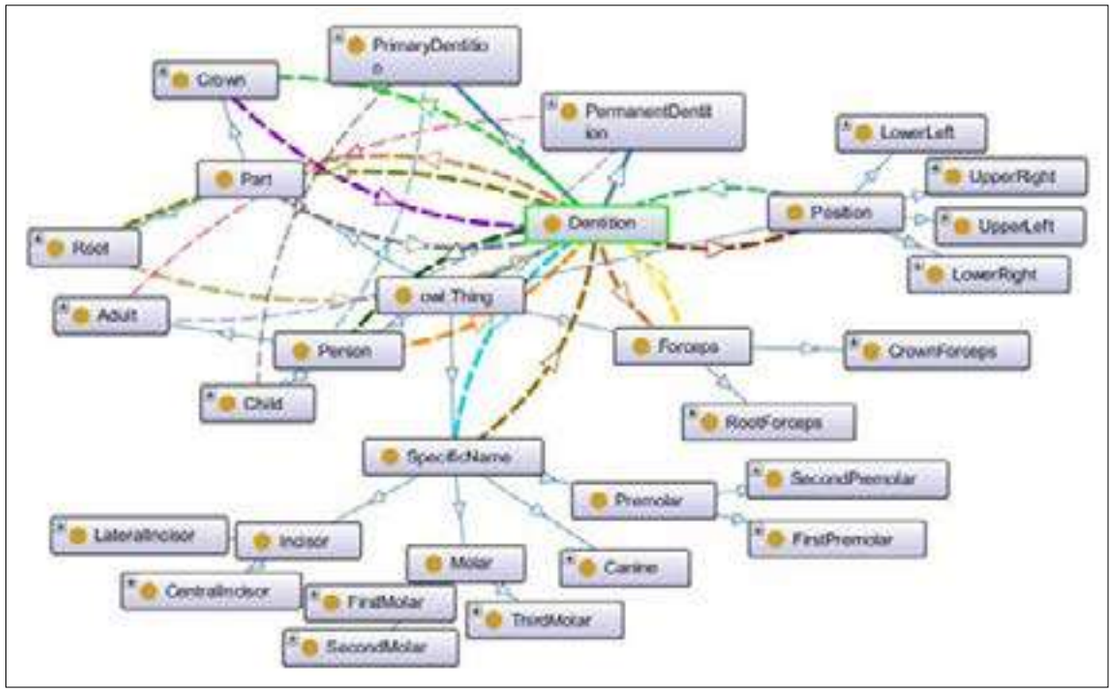


Figure 2. Part of the dental extraction forceps ontology

Table 2. DL query and their answers

Competency questions	DL query	Answers
What forceps are used to pluck T55 dentition?	Forceps and hasUsedToPluck value T55	Instances (2) • ChildUpperMolarCF • ChildUpperRF
What are the dentitions which consist of specific name canine?	Dentition and hasSpecificName value Canine	Instances (8) • T13 • T53 • T23 • T63 • T33 • T73 • T43 • T83
What is the position of T34 dentition?	Position and isPositionOf value T34	Instances (1) • LowerLeft
What is the specific name of T28 dentition?	SpecificName and isSpecificNameOf value T28	Instances (1) • ThirdMolar

In order to avoid the defects when using the ontology, its quality should be verified and validated (Poveda-Villalón et al., 2012). So, the evaluation of the ontology was conducted at the last stage of our methodology which is done by experts in the field ontology and by using inbuilt FaCT++ 1.6.5 reasoner. Further, in order to detect potential pitfalls which will lead to errors in modelling, we used an online ontology evaluator called OOPS! which is available at http://oops.linkeddata.es/ (Poveda-Villalón et al., 2012). This evaluator evaluates the developed ontology in five aspects; (i) human understanding, (ii) logical consistency, (iii) modelling issues, (iv) real-world representation and (v) semantic applications for the developed ontology.

IV. DISCUSSION

Tooth extraction is one of the common surgical procedure in the field of dental, which mainly depends on the knowledge and experience of the dentists. Great importance on knowledge for dental extraction forceps is given in this research. An ontological approach to representing the necessary dental extraction forceps knowledge within the dentists' context was developed.

Designing this type of ontology is not a simple task, because we need to gain vast domain knowledge. In this paper, we have explained how we designed and developed the ontology to organize domain knowledge by meeting particular access requirements effectively. Using this approach, dental extraction forceps ontology to include information needs to be identified for dental extraction activities to be designed.

The validation and evaluation have been done separately. We validated the ontology in terms of accuracy and quality by using the FaCT++ reasoner which is and inbuilt tool in Protégé-OWL Ontology Editor 5.1 and by using web-based tool OOPS!. We evaluated the ontology with the help of ontology expert by examining the deficiencies of the artifacts we used. Based on Ontology experts' responses, comments, and suggestions the ontology was redeveloped.

V. CONCLUSION

Tooth extraction is one of the common surgical procedures in the field of dental, which mainly depends on the knowledge and experience of the dentists. Therefore, in this research, great importance is given to knowledge for dental extraction forceps, and the potential benefits of managing dental extraction forceps knowledge. An ontological approach is needed to represent the necessary dental extraction forceps knowledge within the dentists' context (Vasanthapriyan et al., 2017a). Identification of the suitable dental extraction forceps for the given case is resolved by developing a domain ontology on dental extraction forceps. Our dental extraction forceps ontology not only solves the problem of selection of right dental extraction forceps but also, it provides a valuable knowledge sharing method for dental students and other researchers.

Designing the ontology in the dental domain is not a simple task. The complexities in the dental extraction domain and the need to gain vast domain knowledge made this task tedious. This research presents dental extraction forceps ontology which represents dental extraction forceps domain knowledge. It includes dental extraction forceps concepts, their properties, and their relationships. We strongly believe that our dental extraction forceps ontology can support the dental hospitals, dental students and other active researchers in this field to improve the sharing of knowledge and experiences.

Our future works have two main parts. Firstly we are planning to expand our research to the whole extraction process which includes all the devices used in the extraction process. Secondly, the development of a knowledge management portal for our expanded research in order to disseminate the knowledge on the dental extraction process. Even though there are some works on knowledge sharing such as software testing (Vasanthapriyan et al., 2017b), economics (Yoo and No, 2014) etc., a very little research has been conducted on the dental domain. So our plan is to develop a knowledge management portal in the domain of dental extraction.

ACKNOWLEDGEMENT

We acknowledge the experts and other personnel who were directly and indirectly involved in order to complete the research work successfully.

REFERENCES

ANYANECHI, C. & CHUKWUNEKE, F. 2012. Survey of the reasons for dental extraction in eastern *Nigeria. Annals of medical and health sciences research*, 2, 129-133.

BALAJI, S. 2013. Burried broken extraction instrument fragment. *Annals of maxillofacial surgery*, 3, 93.

CHEN, R.-C., JIANG, H. Q., HUANG, C.-Y. & BAU, C.-T. 2017. Clinical Decision Support System for Diabetes Based on Ontology Reasoning and TOPSIS Analysis. *Journal of Healthcare Engineering*, 2017, 14.

CHOKSI, A. T. & JINWALA, D. C. 2015. A Novel Way to Relate Ontology Classes. *The Scientific World Journal*, 2015, 15.

GARCIA-VALVERDE, T., MU, #XF1, OZ, A., #XE9, ARCAS, F., BUENO-CRESPO, A., #XE9 & CABALLERO, A. 2014. Heart Health Risk Assessment System: A Nonintrusive Proposal Using Ontologies and Expert Rules. *BioMed Research International,* 2014, 12.

GRUBER, T. R. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43, 907-928.

GRÜNINGER, M. & FOX, M. S. 1995. Methodology for the design and evaluation of ontologies.

HEIMANN, W. 1977. Extraction forceps for upper molars. Google Patents.

HU, Y., ZHOU, W., REN, J., DONG, L., WANG, Y., JIN, S. & CHENG, L. 2016. Annotating the Function of the Human Genome with Gene Ontology and Disease Ontology. *BioMed Research International*, 2016, 8.

J. F. CCLYER, L. R. C. P., M.R.C.S. L.D.S. 1986. *Extraction of the Teeth*, London, Claudius Ash & Sons, Limited.

KOLOSOVAS-MACHUCA, E. S., J., F., POZOS-GUILLEN, A. J., CAMPOS-LARA, N. P. & PIERDANT-PEREZ, M. 2016. Pain Measurement through Temperature Changes in Children Undergoing Dental Extractions. *Pain Research and Management*, 2016, 5.

KUZIEMSKY, C. E. & LAU, F. 2010. A four stage approach for ontology-based health information system design. *Artificial Intelligence in Medicine*, 50, 133-148.

PARK, S. & KIM, H.-G. 2006. Dental decision making on missing tooth represented in an ontology and rules. *The Semantic Web–ASWC 2006*, 322-328.

PISANELLI, D. M. 2004. Ontologies. Available: http://www.openclinical.org/ontologies.html [Accessed 22/11/2017].

POVEDA-VILLALÓN, M., SUÁREZ-FIGUEROA, M. & GÓMEZ-PÉREZ, A. 2012. Validating ontologies with oops! *Knowledge Engineering and Knowledge Management*, 267-281.

RAO, R. R., MAKKITHAYA, K. & GUPTA, N. Ontology based semantic representation for Public Health data integration. Contemporary Computing and Informatics (IC3I), 2014 International Conference on, 2014. IEEE, 357-362.

TEVFIK OLUREL, Z. O., REYHAN OLUREL, NURGÜL ÖRNEK, NESRIN BÜYÜKTORTOP GÖKÇJNAR, AND KEMAL ÖRNEK 2014. Branch Retinal Artery Occlusion following Dental Extraction. *Case Reports in Ophthalmological Medicine*, 2014, 3.

VASANTHAPRIYAN, S., TIAN, J. & XIANG, J. An Ontology-Based Knowledge Framework for Software Testing. International Symposium on Knowledge and Systems Sciences, 2017a. Springer, 212-226.

VASANTHAPRIYAN, S., TIAN, J., ZHAO, D., XIONG, S. & XIANG, J. An ontology-based knowledge management system for software testing. The Twenty-Ninth International Conference on Software Engineering and Knowledge Engineering (SEKE), 2017b. 522-525.

WALISADEERA, A. I., GINIGE, A. & WIKRAMANAYAKE, G. N. 2015. User centered ontology for Sri Lankan farmers. *Ecological Informatics*, 26, 140-150.

WALISADEERA, A. I., WIKRAMANAYAKE, G. N. & GINIGE, A. An ontological approach to meet information needs of farmers in Sri Lanka. International Conference on Computational Science and Its Applications, 2013. Springer, 228-240.

WIKIPEDIA. 2017. Dental extraction [Online]. Available: https://en.wikipedia.org/wiki/Dental_extraction [Accessed November 01 2017].

YOO, D. & NO, S. 2014. Ontology-based economics knowledge sharing system. *Expert Systems with Applications*, 41, 1331-1341.

AUTOMATED PREDICTION OF CUSTOMER HOTSPOTS TO TAXI DRIVERS USING CLUSTERING TECHNIQUES AND WEB SCRAPING

SPDT Saranatha[1] and PPNV Kumara
Department of Computer Science, General Sir John Kotelawala Defence University, Sri Lanka
[1]saranatha@yahoo.com

Abstract- Taxi service is one of the most important service in our society. There are many mobile application to customer to book a taxi. But the problem is that applications are not utilized properly to find taxis to customers in a city area or a busy environment because the demand for taxi exceeds the supply. Purpose of this research is to predict the taxi travel demand in a city area. So that mobile applications can utilize properly to guide taxi drivers to give a proper service to customers. The prediction is done clustering the historical data such as time, weather, location using clustering techniques like k-means and DBScan we can cluster the data and cluster hotspots of customers can be found. There are websites that shows details about the upcoming events. In that websites we can find event location, time and other details about the event. So using web scraping techniques we can scrape those data to get that event data. Using those data we can notify the taxi drivers about the nearby events. So they can easily find more customers who are attending to those events quickly. By this method the time and money of both taxi drivers can be saved. So the profit of taxi drivers will be increased.

Keywords- clustering, web scraping, taxi

I. INTRODUCTION

Studying the pattern of people movements gives lot of information about their daily activities and location is one of the most commonly used forms of context. It is generally simple to gather location information, and other pieces of context may be inferred from location, for example, the nearness of other people (Daniel Ashbrook, 2003). People can be provided with many services using this information. So that people can increase the productivity of their day. Food companies, cloth companies, taxi companies need this type of data to find the hotspots of customers. This paper describes more about how taxi companies can improve their profit using these types of data. To study this moving patterns and converting this into statistical format and to see that is there any technical format by which we can analyse his moving behaviour. The study of consumers helps taxi companies to improve their strategies. One reason for driving an empty vehicle is that taxi drivers do not know where potential clients are, abandoning them with no decision however to meander around the city. The goal of this research is to predict the areas with potential demand from contexts and past history.

Global Positioning System (GPS) hardware to collect location data in a simple and reliable manner (Daniel Ashbrook, 2003). Therefore GPS can be used to get the locations of people. Analysing the data on past history, including the time and location passengers got on taxis, provides clues to the demand distribution. Given the contexts of time, location, and weather, relevant records are filtered for further computation (Neema Davis, 2016). And also the other main important thing to concern is extraction of events data from web because events are the places that we can find many customers. And normally customers are not taking their own vehicles to events because of parking problems, some customers drink alcohol in the events and vehicle safety so with these problem some are not taking their vehicle to event so taxi drivers can provide them transporting service if they could find the location and time of the event so extraction of event will give many opportunities to taxi drivers to find customers. Hence the past data can be collected from taxi service providers and using proper data mining tools and extracting events from web the hotspot of the customers can be calculated.

II .LITERATURE REVIEW

Detecting hotspot and studying the behaviour pattern of people is a highly interested area. Due to availability of fast computations there is more access to the technology of hotspot detection. In the paper Hotspot detection and clustering: ways and means (B.Lawson, 2010) it defines what a hotspot is and what is clustering. The most general meaning of clustering and hence clusters is where an intensity threshold or level threshold is used and any area of a map above the threshold counts as a cluster (B.Lawson, 2010). This term level thresholding furthermore, it is generally fluffy definition as there is no necessities for bunches to have neighbour integrity, be a sure shape or measure or to be particular as for this situation no limit prerequisites are characterized. This is fundamentally a type of hotspots clustering where a guide is checked for regions of 'excess level. In general, the distinguishing proof of clusters requires the meaning of cluster location, cluster size, cluster shape and furthermore potentially some measure of least force. For instance, in Figure 1 board A shows a substantial bunch though Figure 1 board B may recommend the presence of little clusters (or no groups by any means). Size is obviously a distinctive factor for this situation.

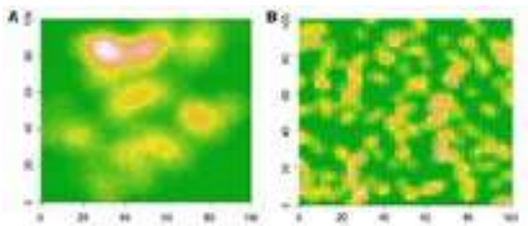


Figure 1. Types of hotspots

In Taiwan each taxi driver operated the business 9.9h a day averagely. But they were wasting about 3.2 h in road without taking any passengers. So (Han-Wen Chang, 2010 january)in there research they have done this to reduce the time of taxi drivers who are roaming for fairs without a proper guide. In their research they analyse the data in past history, including the time, location, and weather. And they have used the clustering methods to find the locations of the customers with a high density or the location of customers who are requesting taxies highly. GPS signals were used to calculate the location of the customer and since the signal isn't precise the records were not indistinguishable but rather spatially near each other so they gathered the adjacent areas into clusters, and that groups additionally mapped to streets or landmarks which covers the greater part of the focuses in the cluster which is showed in figure 1. In the below illustration they show that request records (the plus-sign points) undoubtedly form three main clusters(with few outliers). In the intersection of road A and road F the upper right most cluster is formed. And in road E between road B and C the lower cluster is formed. So that is how they illustrate the formation of clusters in road. They have used three clustering algorithms k-means, x-means, hierarchical clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN)(Martin Ester, 1996)). Also, they expressed distinctive clustering strategies had diverse performances on various sort of data distributions with the goal that it was difficult to choose one as the best.

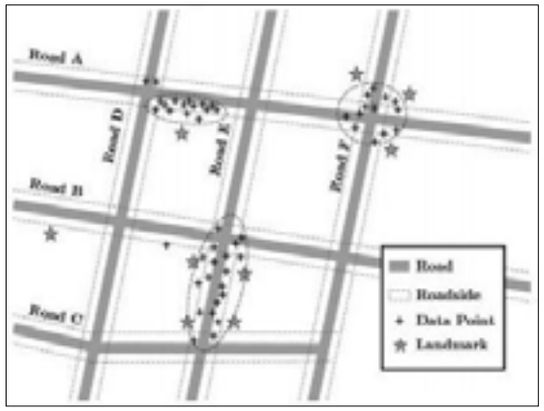


Figure 2. Mapping clusters to road

In the research a context aware taxi demand hotspots prediction is done using data mining techniques they only consider about finding clusters not the optimum clusters, clusters which are more profitable to taxi drivers. And they have included in their future works to consider the events in that area (social event) that a taxi driver can find more customers.

(P. X. Zhao, 2015) has done a research for detecting hotspot from taxi trajectory data using spatial cluster analysis. A method of trajectory clustering based on decision graph and data field. Contrasted and normal clustering techniques, it can naturally find out parameter as opposed to doing that by encounter and is reasonable to direction clustering. Furthermore, they apply it to direction based urban hotspot disclosure in Wuhan City, China. .Circulation and progression of the hotspots are investigated by utilizing taxi direction information regarding occasion, weekday and weekend. However, like the majority of the current clustering algorithms, the proposed strategy just thinks about spatial data, estimating similarity of points with distance between them. A trajectory data point is a trace generated by a moving object in geographical space usually represented as by series of chronological ordered points. In their paper they have mentioned that existing clustering algorithms like k-means and DBSCAN for hotspot detection discovery have some difficulties in meeting requirement of trajectory data for heterogeneous spatial distribution. Though they predict hotspot as well as traffic hotspot they didn't consider about finding the optimum clusters which are profitable to taxi drivers and finding the clusters considering the traffic data. And also they didn't include about considering event which are highly affects to the prediction of hotspots.

In the paper Predicting Taxi-Passenger Demand using Streaming Data(Luis Moreira-Matias, n.d.) they exhibited a novel use of time arrangement of time series forecasting techniques to enhance the cab driver versatility insight. They took signal emitted by 441 taxi from an organization working Porto, Portugal by transforming the GPS and occasion signals into time series to use firstly as learning base to their model and secondly as a streaming test framework. Accordingly, their model could foresee the taxi-traveller request at every single one of the 63 taxi remains at 30 minute time span interims. Furthermore, they said that the model made by them showed a more than satisfactory performance, accurately forecasting 506873 tested services with an aggregated error measure lower than 26%. That system ready to mine both the periodicity and seasonality of the traveller request, refreshing itself frequently, It at the same time utilizes long-term, midterm and short term authentic information as a learning base and it takes focal points of the pervasive attributes of a taxi network, amassing the experience and the learning everything being equal/drivers while they for the most part utilize only their own. In this research they didn't consider the weather data to their prediction. Weather is a highly effectible matter for the hotspots prediction and also they didn't consider about finding optimum clusters and also consider social event data.

In some countries the traffic changes from area to area very rapidly so a one model so unified model which can apply to countries with the same traffic in all areas cannot be apply to countries with different traffic conditions in different areas. So Neema Davis, Gaurav Raina and Krishna Jagannathan has done a research A Multi-Level Clustering Approach for Forecasting Taxi Travel Demand(Neema Davis.,2016) to make different models to different areas with different traffic conditions. They have used GPS data to find the location of the customer. Also, a time series model is based on the suspicion that the present and the future request would have some relationship with the past request, and represented as the function past information. So these are the models that they have used, moving average, exponential smoothing, i.e, Holt-Winters (HW) model, Seasonal Naive, FFT, STL decomposition, ARIMA, TBATS, linear regression and state space ETS model. So they have calculated which model fits to which area best. To do that they have used baseline model which is an averaging model so that a model is retained if it performs better than the baseline, else it is discarded. And they used a multi-level clustering technique to improve the performance of the model by 20%. In this research the importance is that they have made different models to different areas in the country but they didn't consider the weather data and also the social evet data. In here though they predict the hotspot they didn't consider about finding the optimum hotspot.

Mining hotspot are more usable to the new taxi drivers who are doing taxi service as a start-up. Because like typical taxi driver a new taxi driver doesn't know about the customer he don't have good experience about customer behaviour patterns for him prediction of cluster centres is very useful. There are taxi drivers who do taxi service as a part time job. So for taxi drivers like that it takes time to learn about customers. For by using a taxi services application those taxi drivers can find hotspots and can go that places and can provide good service rather than roaming around the city uselessly. And for the typical taxi drivers mining and predicting hotpot is not very useful but if they can know the optimum hotspots then that is useful for them.

System	Use time, location for the prediction	Use weather data for the prediction	Use clustering algorithms	Use social event data	Use time series analysis	Use decision graph	Find the optimum cluster
A	✓	✓	✓	✗	✗	✗	✗
B	✓	✗	✓	✗	✗	✓	✗
C	✓	✗	✗	✗	✓	✗	✗
D	✓	✗	✗	✗	✓	✗	✗
Our system	✓	✓	✓	✓	✗	✗	✓

Figure 3. summary of review

A - A context aware taxi demand hotspots prediction.

B - Detecting hotspot from taxi trajectory data using spatial cluster analysis.

C - Predicting Taxi-Passenger Demand using Streaming Data.

D - A Multi-Level Clustering Approach for Forecasting Taxi Travel Demand.

III. METHODOLOGY

We have used Kaggle dataset of New York taxi trip duration and also kaggle New York hourly weather data set and we combined those two datasets. So we can have both taxi trip data and weather data. And we used Scikit-learn library in python to do the clustering. To prediction of hotspot we have done three things.

1. Clustering the dataset considering date, time, weather and location.
2. Extraction of social events data from evet notifying web sites.
3. Prediction of hotspots and notify about events to taxi drivers.

A. K-means algorithm

In this technique the number of cluster (k) is predefined preceding examination and afterward the choice of the underlying centroids will be made arbitrarily and it took after by iterative procedure of doling out every datum point to its closest centroid. This procedure will continue rehashing until the point when merging criteria met.

In this algorithm first we have to define the number of clusters. This algorithm work as follows: In the following illustration there are 5 data points and it takes k=2. And in the figure 4 it randomly assigns each data point to a cluster.

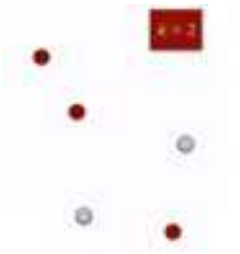


Figure 4.

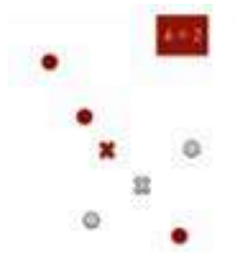


Figure 5.

Then the centroid of each cluster have to be compute so in figure 5 the centroid of data points in the red cluster is shown using red cross and those in grey cluster using grey cross.

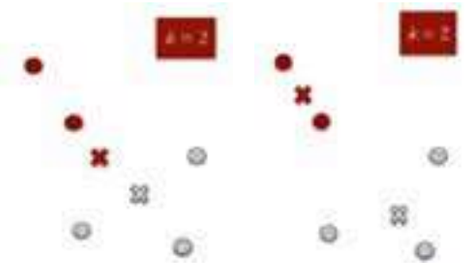


Figure 6. ***Figure 7.***

Figures 4,56,7 Data points(saurau,2016)

Then in figure 6 each data point is re-assign to the closest cluster and again new cluster centroids have to be re-compute. And it is shown in figure 7

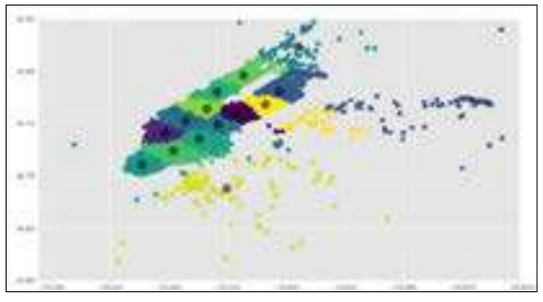


Figure 8, output of K-means

To our dataset when we apply K-means algorithm the output is shown above figure 8. We have used silhouette score to gauge accuracy and it is an esteem is a measure of how comparative a object is to its own cluster (cohesion)

contrasted with different clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and inadequately coordinated to neighbouring clusters. The silhouette score of the above cluster data is 0.4412.

In the above algorithm the problem is that the number of clusters should be predefined and it is sensitive to noise data. And when we reducing the cluster number the silhouette score get increase to some extent but the problem is each time we have to define the number of clusters.

B. DBSACN algorithm

In k-mean algorithm we can't discover the noise properly. DBSCAN (Density Based Spatial Clustering of Applications with Noise) which is designed to discover the clusters and the noise in a spatial database(Martin Ester, 1996)).And this algorithm work as follows:

There are two parameters:

- Eps – Maximum radius of the neighborhood
- MinPts- Minimum number of points in the Eps-neighborhood of a point.

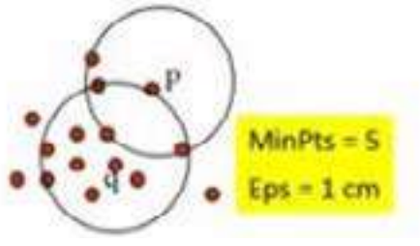


Figure 9. Data points(saurau,2016)

A point should be arbitrarily selected(P) and retrieve all points from p w.r.t. Eps and MinPts and if p is a core point cluster is formed and if p is border, no points are density-reachable from p, and DBSCAN visits the next point of the database. And it continues the process until all of the points processed.

When we apply DBSCAN algorithm to the dataset we get a clustering shown in the figure 8. And the silhouette score is 0.01970. When we applied as shown in the figure the noise data do not get into clusters. But problem is we need cluster centre points so in here it do not gives directly cluster centre points.

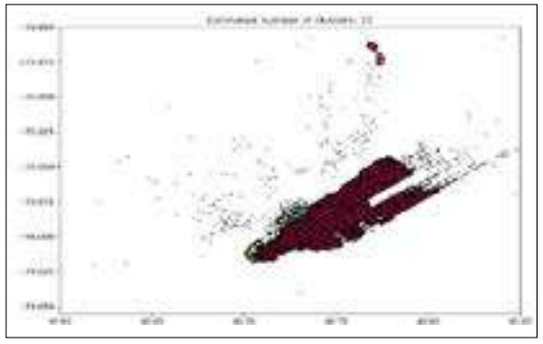


Figure 10. Output of DBSCAN

C. DBbkmeans Algorithm

In the paper A Novel Density based improved k-means Clustering algorithm(K. Mumtaz, 2010) they propose a new algorithm which overcomes the draw backs of DBSCAN and K-means clustering algorithms. In this research they have combined the DBSCAN and K-means algorithms. They have first run the DBSCAN algorithm and then they have find the number of clusters and then using the cluster value they run the K-mean algorithm. So in our project we used this combined algorithm to find the clusters.

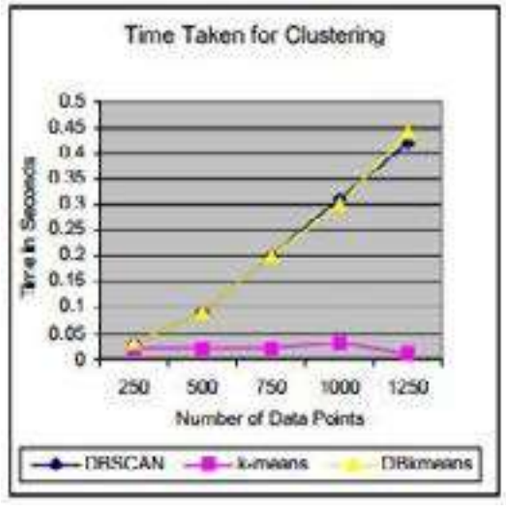


Figure 11. Time taken for clustering(K. Mumtaz,2010)

As shown in the above figure 11 time taken for clustering is increasing when number of records are high and that is disadvantage but in our project since we are updating the hotspot hourly it will not be big issue.

Algorithm	Silhouette Score
K-means	0.44012
DBSCAN	0.01970
Dbkmeans	0.48403

Figure 12. silhouette scores

As mentioned in the above table the silhouette score for each algorithm when we done the clustering using different algorithms are shown. So according to that the Dbkmeans shows the highest score. Because of that using that algorithm to our system will give more accurate results.

After calculating clusters the next task is to select the optimum clusters. To calculate the optimum cluster mean that the most profitable cluster to the taxi driver. To do that we first calculate distance between pickup and drop-off location using google maps API and then we multiply that distance by charge(cost)taken per unit distance. So we do that for all the data points in a cluster and add all the values and take the total. Then that total value is divided by the total number of taxi rides in that cluster. The final answer is the average income of taxi driver who use that cluster.

- Distance between pickup and drop-off location in data point in a cluster – distance(i,(x,y)) -> i- number
- of the cluster, (x,y) – coordinate of the data point in the cluster.
- Total Distance of the all the data points in the cluster Σ distance(I,(x,y)).
- Total income of the cluster=Total Distance × Charge per unit distance.
- Average income of a taxi driver = Total income ÷ total number of points in the cluster

So the cluster with the highest average income of the taxi driver will be the optimum cluster.

D. Extraction of events from web

An occasion happens at a specific area, has a begin date and time, and a title or portrayal. In other words to be valuable to a client, an occasion must have the capacity to answer the inquiries: What?,When?, and Where?(John Foley, 2015). In the paper learning to extract local events from web (John Foley, 2015). So they have focused on the identification and extraction of events on the open internet and recommend events that users might want to attend. So this method is very useful to taxi drivers because they can find more customers in an event area. So the taxi service providers can use this method to notify the location and the time of the event to taxi drivers.

Web scraping is also known as web data extraction, web data extraction, web harvesting, and screen scraping. And it is great technique of extracting unstructured data to a structured format. So that structured data can be stored in databases, spread sheets, XML files. Information like online price comparison, context scraping, weather data monitoring, extract offers and discounts, and extract information from job notification web sites, collect government data and market data are the data that can scrap using web scraping.

E. Web scraping techniques

There are several methods and ways that we can do web scraping. And each one has its own advantages and disadvantages

Computer vision web page analysers, they can analyse the web page like human using machine learning and image processing techniques and find the important unstructured data and store them in structured format. Example is Diffbot. And this technique uses high computation power because it using both machine leaning and image processing techniques.

In our research we have used web scraping using python and the BeautifulSoup library and we used it because the simplicity and easy to handle. But the problem in that method is to we have to customize the web scraping code to site to site because in that method what we do is we extract text from HTML tags so tag arrangement is differ from site to site. So by that method we can find the location, time and name of the event. And store that data

in the database so that the taxi service provider can use that data to notify the taxi drivers.

F. Client and server side

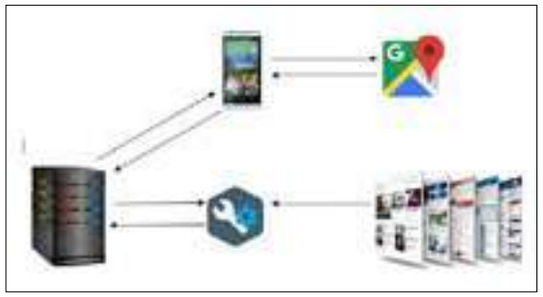


Figure 13. Data flow

The entire system is designed to be implemented in a client server architectural design. The clustering model is stored in the server. And when the taxi driver request for hotspot the server run the clustering model considering location, time, and weather condition. And also considering the location it finds the event details of the nearby locations using the web scraped data. And it process all data together and send it the taxi driver. As shown in the above data flow diagram first the taxi mobile application should pass the location, time, weather data to the service providers server. Then the server process the data and start clustering considering location, time, weather condition and it output the hotspot data to taxi application. And the service providers server run the web scraping application by passing location and time data to the web scraping application so the application has stored unstructured scraped data in a structured format and that application use that data and pass the event details considering location and time given by the server. And the server pass the that event details data to the taxi application and taxi application uses google map API to find the location send by the service providers server and then the hotspot location and event details are shown in the taxi application.

IV. CONCLUSION & FUTURE WORK

Automated prediction of customer hotspot system for taxi drivers system is introduced to overcome from the problem in a city area the demand for taxi exceeds the supply. This system cluster the historical taxi booking data like date, time location and the weather condition at time of the booking to calculate customer hotspots. And it collect the data from web site to find about events and notify the taxi driver about the customer hotspots. And it gives the most optimum cluster that taxi drivers should use it order to have higher income. As future work traffic data also should be considered when we finding the optimum hotspots to taxi drivers and finding most accurate and fast clustering techniques to cluster the historical data.

ACKNOWLEDGMENT

We would like to thank all the people who helped us in doing a lot of Research and we came to know about so many new things and we are really thankful to them.

REFERENCES

Anja Struyf, M. H. ,. J. R., 1997. Clustering in an Object-OrientedEnvironment. Journal of statical sofware, Issue 4.

B.Lawson, A., 2010. Hotspot detection and clustering: ways and means. p. 15.

Chang, H.-W., 2010. Context-aware taxi demand hotspots. International Journal of Business Intelligence and Data Mining , p. 7.

Daniel Ashbrook, T. S., 2003. Using GPS to Learn Significant Locations and Predict Movement Across. Personal and Ubiquitous Computing , 7(5), pp. 275-286.

Han-Wen Chang, J. Y.-j. H., 2010 january. Context-aware taxi demad hotspots prediction. Internation Journal of Business Intelligence and Data Mining , p. 17.

Hassan Sayyadi, M. H. M., 2009. Event Detection and Tracking in Social Streams. Proceedings of the Third International ICWSM Conference .

John Foley, M. B. J., 2015. Learning to Extract Local Events from the Web.

K. Mumtaz, D. K. D., 2010. A Novel Density based improved k-means. (IJCSE) International Journal on Computer Science and Engineering, Volume 02, pp. 213-218.

Luis Moreira-Matias, J. G. F. D., n.d. Predicting Taxi-Passenger Demand using Streaming Data.

Martin Ester, H.-P. K. ,. J. S. ,. X. X., 1996). A density-based algorithm for discovering clusters in large spatial databases with noise.

Neema Davis, G. R. a. K. J., 2016. A Multi-Level Clustering Approach for Forecasting Taxi Travel Demand. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).

P. X. Zhao, K., 2015. DETECTING HOTSPOTS FROM TAXI TRAJECTORY DATA USING SPATIAL. International workshop for spatiotemporal computing , p. 5.

Wang S, G. W. L. D. e. a., 2011. Data field for hierarchical. Journal of Data Warehousing and Mining (IJDWM).

SaURAV KAUSHIK , NOVEMBER 3, 2016(https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering)

IMAGE PROCESSING BASED AUTOMATIC PELICON CROSSING SYSTEM

KGL Chathumini[1], BTGS Kumara[2], and HACS Hapuarachchi[3]

[1,2]Department of Computing & Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.

[3]Department of Sports Sciences & Physical Education, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.

[1]*lohara.chathumini@gmail.com*

Abstract- Traffic congestion and pedestrian accidents are two major issues that the Sri Lankan society face today. Many social, economic and environmental problems increase due to the two issues. Lack of effective Pedestrian Light Controlled (PELICON) crossing system in Sri Lanka is one of the reasons for road traffic congestion problems. Thus in this paper, we propose Image Processing Based Automatic PELICON System. The proposed system first takes an image input to the system using CCTV camera. Then, it identifies the particular color ranges using color detection algorithms. Next, noises are removed using filters. Finally, the system uses background subtraction and contour analysis algorithms to identify pedestrian object contours. The system then calculates the number of pedestrians. If the pedestrian object count is greater than a given threshold value or if the pedestrian waiting time is exceeded, then the system shows green signal to pedestrians and red signal to vehicles. Empirical study of our prototyping system has proved the effectiveness of our pedestrian detection approach. Further, we conducted a questionnaire survey to check the suitability of this system to the Sri Lankan society. Randomly selected people were taken to this study as the sample population. According to the study results, the system will minimize road traffic congestion and other negative impacts of the traffic congestion.

Keywords- Color detection, Background Subtraction, Contour analysis

I. INTRODUCTION

Colombo is the largest city and major economic centre in Sri Lanka with a metropolitan area of 5.6 million inhabitants, who are expected to increase up to 8.4 million by 2030. Main domestic transportation mode in Colombo, in Sri Lanka is land transportation. But, supply road space for the demand of land transportation is low. The major urban transport issues in Sri Lanka are traffic congestion, environmental pollution, increase of road traffic accidents, poor public transport system and weakness of road network. There are several factors that cause these urban transportation issues. Some of them are: high population of people in Colombo, high population of vehicles in Colombo, increase of road accidents, high pedestrian deaths, etc. These outputs interact within the prevailing social, economic and environmental aspects, and they produce negative outcomes.

Traffic congestion, pedestrian accidents and other crummy impacts of traffic congestion of Sri Lanka are the noticeable problems in Colombo city during peak periods or rush hour (Kumarage, A.S., 2004). Both vehicles and pedestrians represent as most dominant users of road space.

There are 10 major entry corridors to Colombo city, approximately 200,000 vehicles enter to the city daily

and it takes 750,000 people (Weerawardana, 2011). The balance of around 175,000 private vehicles carry 1 or 2 passengers each. Presently around 15% of the road space is utilized for bus transport even though it transports 62% of road passengers. On the other hand, 65% of the road space is used by private and hired vehicles which is sum total carry only 38% of passengers. However, there is no proper way to handle both of the pedestrians and vehicles.

Withal, traffic problem aggrandise social economic problem as well. Some of them are, loss of competitiveness in Stock market, inflationary price due to unproductive resources, migration of businesses to peripheral areas resulting in spreading of the city and increase in transport and other utility service costs, which are in turn impact above two problems, increases of environmental pollution, social discontent and loss of political goodwill, etc.

Each year over 40,887 road accidents are occurred in Sri Lanka, causing on average six fatalities every day. Nearly 9000 car and van accidents are occurred annually with over 5400 accidents are caused due to speeding. Over 740 pedestrians are died every year, recording two pedestrian deaths per day. Of the total number of road accidents that occur each year, over 2471 result in accidental deaths. Several hundred are left seriously injured, some with lifelong consequences. There are over 1000 road accident per-week with 5 or 6 people have been killed every day. The economic cost of accidents has been valued over Rs 10,000 million annually (Kumarage, Wickramasinghe, & Jayaratne, 2003).

Further, Most of the road-accidents are occurred while the pedestrian crossing the road. Pedestrian fatalities in Sri Lanka account for 40% all road deaths. In Colombo district this is high up to 70% and it represents pedestrian constitute 39% (Fernando, M., 2017). According to the World Health Organization, without any sustained action, road traffic crashes are predicted to become the seventh leading cause of death by 2030.

A. Introduction to PELICON

The PELCON Crossing is standard for PEdestrian LIght CONtrolled Crossing. It is used pair of poles with standard traffic light system and in between that the road is on-going. It is facing oncoming traffic, it has a push button on the pole that is for pedestrian who wants to cross the road to press and it uses two illuminated and coloured pictograms facing the pedestrian from across the road. It uses red, green and yellow signal as usual.

B.Research Problem

For example, some roads like Galle road have several button clicking PELICON systems. Sometimes, one vehicle may have to stop each and every PELICON Crossings while they ride from Ratmalana to Fort using Galle road. Because, this kind of PELICON Crossing systems are responding to single pedestrian either group of pedestrians. This system response evolves high traffic congestion.

The use of PELICON crossing system in an effective way is for more reliable solution to minimize road traffic, road accident and other social, economic and environmental problems of traffic congestion. In this research paper we proposed a system using image processing techniques that can be used to PELICON crossing system and questioner survey to check the suitability of this system for Sri Lankan society.

Main requirements to the system are taken by Colombo Municipal Council's Engineering Division. CMC's Traffic Division is already using image processing technique in their newly developed PELICON Crossing systems. But, they do not considering about the number of pedestrians.

C. Aims and Objectives

The aim of this research is to minimize the traffic jam and pedestrian accident of the Colombo city. Furthermore, minimize environment pollution, flue wastages and all other negative impacts of traffic congestion.

Objectives of this research are: to analyse the existing research work and identify the issues, to develop image processing algorithm to identify the number of pedestrians, to write the algorithm to control the traffic light system, to evaluate the system with existing system and to do a systematic survey, whether this system is suitable for the Sri Lanka society.

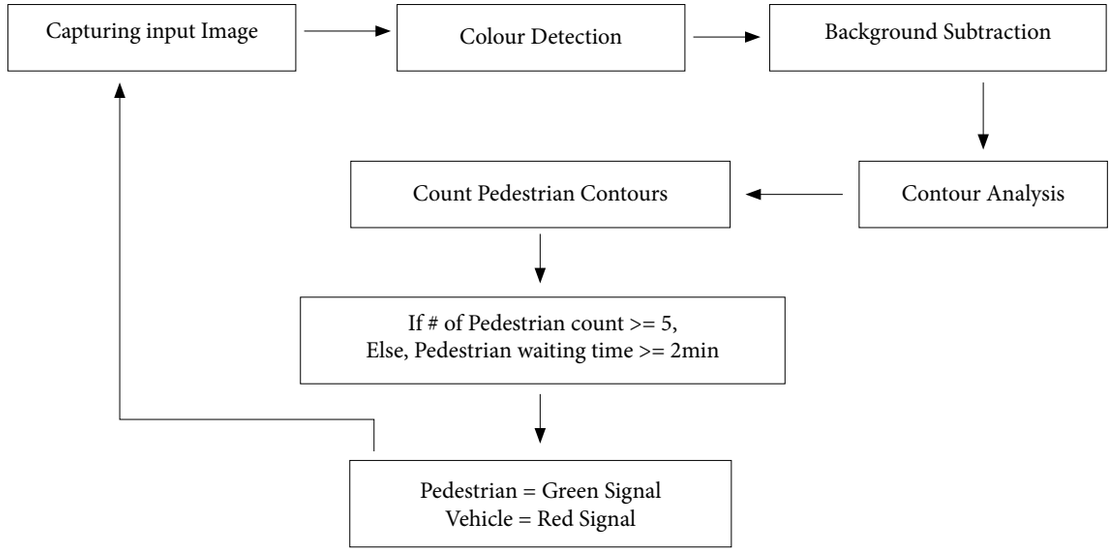


Figure 1. Proposed Approach

II. METHODOLOGY AND EXPERIMENTAL DESIGN

A. Proposed Approach

Figure 1 is the architecture of the proposed approach. Input image to the system taken from CCTV camera (Jaison A., Varghese E., Gopika K.G., and Krishnadas J., 2018) which is located behind of the traffic light pole. CCTV video runs 24 frames per second and one of the image is been taken as an input image to the system with specific time interval. Then, the system is used three major steps to produce the output; color detection, background subtraction and contour analysis.

The first step of the system is the use of color detection algorithm. Then, the system used background subtraction to remove the noises and tune the image. Finally, contour analysis used to analyze the contours and analysed contours are matching with template contours, those are already stored in the system in initiation stage. If the image contours matches with the template contours, the system identifies the contour as a pedestrian object contour. Then, the identified pedestrian object contour count is taken to a variable. If the variable value is greater than 5 or pedestrian waiting time exceeds, then the system shows green signal to pedestrian and red signal to vehicles.

We used C# programming language and Visual Studio 2013 with OpenCV wrapper to implement the image processing module and logical part of the proposed system.

1) **Color Detection:** The colour range is needed to be set in the system to identify all the coloured pixels that have been available on the image or the input image. The color detection used to identify the particular color ranges of the input image.

HSV color model is used in color detection. Because, HSV color mode is performed well in color detection than RGB color mode. The color is manually selected in color selector. First, the input image stored in RGB color mode image typed variable and then, it converted into HSV color mode image variable. The image is then used to find out the areas with a particular color range, where it help to create color recognition. According to the amount of hue saturation and value that system has set the maximum and the minimum color range detected in the system

The users can clearly see the object which is with the selected colours in the triangle; it represents the areas where the color is present in the image. The triangle is being controlled by manipulating the HSV colour amount. The needed colour range is being chosen by use

of the colour palate as shown in fig 2. The system provided default HSV values. But using this interface, which is shown in fig. 2 the user, facilitates to change the HSV values for particular image.

Figure 2. Selecting and specifying the required colour range by the colour range finder with default HSV values

2) **Background Subtraction:** After successful detection of the colour, the system sends the color detected image to the background subtraction step. The system background subtraction is done to remove the background and highlight the foreground. The system is used three algorithms for background subtraction step. First, the system starts checking pixel by pixel of the input pixels of the image; flood fill method is used to fill the areas that are not belonging to the particular region. The other areas that do belong to the specifically needed regions are flooded, and the non-belonging areas are left away. In the system is used Gaussian smoothing algorithm to remove the noises of the image. By the use of eroding method on the image, the system identifies the particular color areas of the image; it thinner the bright areas and bigger the dark zones. The use of these three algorithms for the color detected original image is results in an image with foreground only. This result of this step is displayed in the system interface. The background subtracted image/ foreground extracted image is focused the threshold to turn the image into a binary image. It enhances the boundaries of the edges that are presented on the foreground.

Figure 3. Background subtraction of an image

3) **Contour Analysis:** First the system captured an image from CCTV video. Then, identified and extracted objects that are closer to pedestrians using colour differences. After that, the system will checked the identified pedestrian features using contour analysis. Contour analysis use to solve the main problems of the pattern recognition (Liu S., Luo Y. and Yang S., 2007). The contour is an outline of the object, a population of pixels and extrication from background. In this contour is encrypted by the sequence consisting of complex numbers. Opening point of contour is a fixed pixel. Then, it scanned clockwise and each vector of the offset is noted by a complex number (a+bi). Here, the point of the x-axis denoted by "a" and the point offset of the y-axis is denoted by "b". The system used two basic/major project libraries to analyse the contours. That is ContourAnalysis project library and ContourAnalysisProcessing project library.

The ContourAnalysis project library is used to implement basic operations for contours; contour creation, equalization, evaluation of ICF (Intercorrelation Function) and ACF (Autocorrelation Function), comparison, searching of base templates, etc. The methods called ICF and ACF are used to search contour similarities of two contours. ACF does not depend on a choice of opening point of the contour. ACF is shape descriptor of contour and reduction of time of comparing using apply wavelet convolution of ACF. The minimum norm of ICF is measuring of the similarity of two contours.

The project library ContourAnalysis consist with four classes; Contour class, Template class, TemplateFinder class and FoundTemplateDesc class. The Contour class contains the basic operations of contours. The Template class is used to create base templates. The TemplateFinder class implements the fast searching of the template for the given contour. The FoundTemplateDecs class contains similarity rate angles of rotation and scale of contour relative to templates.

In initiation stage the system, template human object contours have to store on the system. In operation stage, the identified contours of the input image are search and match with the template contours. Then, the matching contours are identified as human object contours. Following fig. 4 illustrates the sample of contour analysis and contour selection processes. White color outline is the source vectors and red color outline is the equalized vector.

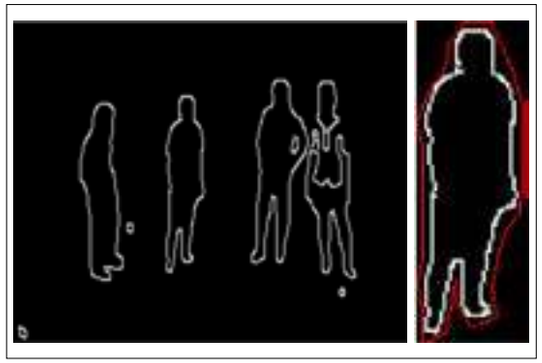


Figure 4. Initial handling of the selection of contours of the image and extracted contours of the image

The identified human object contours are highlighted using green color rectangle. Figure 5 shows sample output of pedestrian object identification and human object counting process.

Figure 5. Detected pedestrian object contours and pedestrian object contour count

Finally, the identified human object contour count taken to a variable and if the variable value is greater than given threshold value or pedestrian waiting time is greater than 2 minutes, the system shows green signal to pedestrian and red signal to vehicles.

B. Survey Analysis

When considering the survey, the sample population to the survey is taken from general public and the sample size is 100. We have considered some factors when selecting the sample; Nationality should be Sri Lankan, Race should not be considered, should be above 15 years and the majority in the sample population should be male.

The questioner was a Google document and it sent to the population using Internet via social media. The first 100 responses are taken to the study as the sample population. These 100 responses are representing different gender, 13 different job sectors and different age groups. Also both experienced and non-experienced with PELICON crossing system and driving a vehicle are participated to the survey.

The questioner consists with four main categories of questions; identification of population, identification of existing system issues and plus points in the view of both drivers and pedestrians and final category consists the questions about the proposed approach. According to the obtained responses, the answers were ranked due to provided values; 'Strongly agree'=5, 'Agree'=4, 'Neutral'=3, 'Disagree'=2 and 'Strongly disagree'=1. Then manually calculated the mean, median, mode and total of strongly agree and agree percentages values of the each question. Finally, we have made some decisions based on the median value.

III. RESULTS AND DISCUSSION

A. Proposed Approach

After doing several analysis and implementations, the result of this proposed approach has been established as an executable system. The proposed system is mainly a computer based system. The system use CCTV camera to capture the image to the system input. Then, use color detection algorithms to identify the particular color ranges. When doing background subtraction, the noises and the background of the image is been removed and highlighted the image foreground. Then, foreground image's contours are identified and analyzed using several algorithms. The identified contours are matched with template human object contours that are already stored in the system. This template pedestrian contours need to store at the initiation stage of the image. The matching contours of the foreground input image are identified as the human object contours and then, the system calculates the human object contour count. Finally, if the object count is greater than given threshold value or pedestrian waiting time is exceed, the system automatically shows green signal to pedestrians and red signal to vehicles.

The input image, background subtracted foreground image and contour analyzed image are shown in the system interface fih.3 and fig. 4. Identified human object contours are highlighted using green color rectangle; which you can see in the fig. 5 with matching percentage and identified pedestrian object count. The template contour creation, storing and deleting fractures also

facilitated in the main interface. Figure 6 illustrates the proposed approach interface.

Rather than that, this main interface facilitated to change the HSV color mode values also.

Figure 6. Proposed approach interface

B. Survey Analysis

When considering the survey, the first category of questions is for identifying the population. The responded people are representing different genders. The male responses are higher than female responses. Also, they are representing different age groups. The following fig 7 represents the summary of received responses for first two questions.

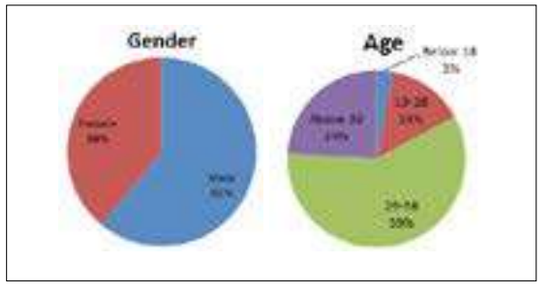


Figure 7. Results of question 1 and 2

Additionally, they are representing different job sectors you can see it in fig 8.

Also, they have different pattern in crossing the road. The question number 4 aimed to identify the pattern of crossing road. Following fig 9 illustrate the results of the question number 4 in the first category.

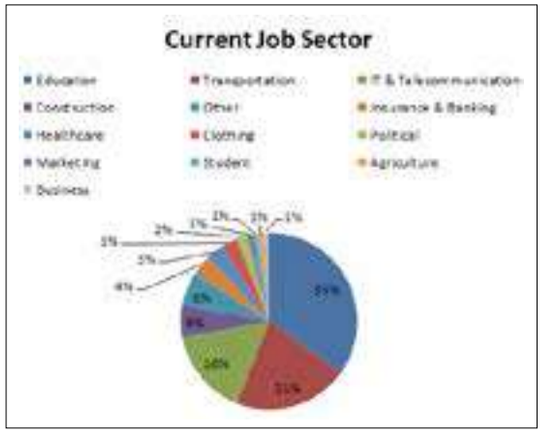


Figure 8. Result of question number 3

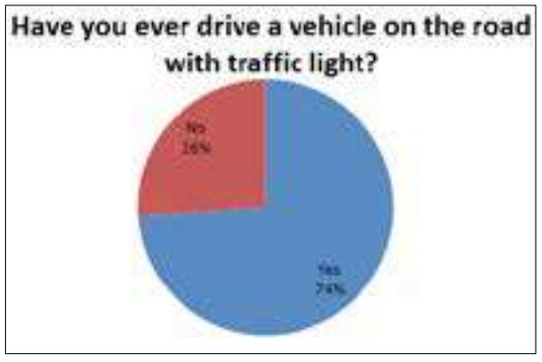


Figure 9. Results of question number 4

Many people in the sample population were used PELICON Crossing System and many people were drive a vehicle on the roads with PELICON Crossing systems at heavy traffic hours. The next figure, fig 10 illustrates it. Therefore, the sample population fair enough for the survey.

The nest two categories; category 2 (as a pedestrian) and category 3 (as a driver) are about existing system findings. Most of the questions in the questioner were received symmetric distribution as the result. Then, we were calculated mean, median, mode and total of 'Strongly agree' and 'Agree' percentages. Several decisions made, by comparing these values with each other.

Final category, category number 4 is about identifying proposed system suitability to the Sri Lankan society.

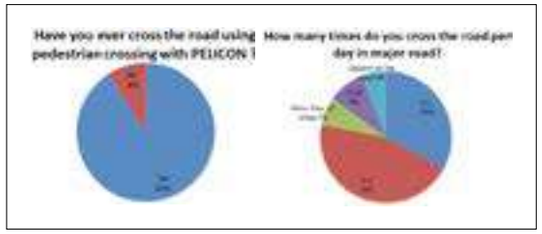


Figure 10. Results of question 5 and 6

Several decisions are made with these question results in the conclusion section.

IV. CONCLUSION

The proposed approach is about effective traffic light controlling system using image processing. As we mentioned, the input image for the system captured using CCTV camera. After several analyses, the system detects human contours and takes the count of human objects. This proposed system suitable for medium and less pedestrian volume areas rather than heavy pedestrian volume areas. Because, the system is needs a very clear input image for processing. But, in generally the pedestrians who are waiting for cross the road anchored as a bunch. So, it is difficult to capture clear CCTV image. We have to notice that, the system identifies dark colors more accurately than light colors. Additionally, at the night time mostly the camera will provide black and white video. This may casus in hard identifying of the human objects at night time. So it is better to use this system at day time. Instead of that we did a questioner survey analysis with randomly selected 100 sized sample population to check the suitability of this proposed system to Sri Lankan society. According to the responses obtained, we have made the following outcomes as a summery. Majority of the Sri Lankans think that; 'Traffic congestion' in Sri Lanka is getting increases day by day and it affects to the Sri Lankan economy badly. This proposed system will be helpful to minimize traffic congestion. The driver will feel comfortable in the cause of using this system in heavy traffic hours/peak hours. 'Pedestrian safety' is a responsibility of pedestrian themselves. This system might not help to minimize pedestrian accidents while pedestrian crossing road. As a future work, we hope to develop vehicle identification algorithm, vehicle speed measuring algorithm and vehicle count calculation algorithms to improve of this system more.

ACKNOWLEDGEMENT

I would like to express my worm thanks to the Dean of the Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka and Vice Chancellor of The Sabaragamuwa University of Sri Lanka.

Further, I would like to express my deepest thank and gratitude to everyone who supported me through this research specially Mr. Palitha Nanayakkara, Commissioner of local government, Western Provincial Council.

REFERENCES

Fernando, M. (2017, June). Retrieved from Digital Library, University of Moratuwa:http://dl.lib.mrt.ac.lk/handle/123/1010

Jaison A., Varghese E., Gopika K.G., and Krishnadas J. (2018, March). Time to cross – traffic light control system using image processing. *International Research Journal of Engineering and Technology (IRJET)*, 5(3), 918-921.

Kumarage, A. S., Wickramasinghe, S. M., & Jayaratne, M. D. (2003). *Analysis of Road Accident in Sri Lanka*. University of Moratuwa Sri Lanka.

Kumarage, A.S. (2004). Urbon traffic congestion: the problems & solutions. *The Economic Review*.

Liu S., Luo Y. and Yang S. (2007). Shape-based pedestrian detection in infrared images. *Journal of information science and engineering, 23*, 271-283.

Weerawardana, W. J. (2011). *Reduction of Traffic Congestion in Colombo City by Improving Public Bus.* University of Moratuwa Sri Lanka.

A NOVEL ELLIPTIC CURVE BASED MULTI-KEY ENCRYPTION METHOD FOR MULTICASTING SINGLE CONTENT WITH ACCESS CONTROL

TMKK Jinasena[1], RGN Meegama[2], and RB Marasinghe[3]
[1,2]Department of Computer Science, University of Sri Jayewardenepura,
[3]Department of Medical Education and Health, University of Sri Jayewardenepura,
[1]*kasunkosala@yahoo.com*

Abstract- The most remarkable invention in the history of cryptography is the invention of public key encryption in the 1970s. It enables users to have a single encryption on a pair of two unique keys. As a result, the way of delivering many security services has changed drastically. Moreover, it introduced new features such as non-repudiation to the cryptographic world. However, in this paper, we present a novel elliptic curve based multi key encryption method to facilitate a single encryption for multiple users where the resulting encrypted content can be multicast to them with access control. Initially, we establish an elliptic curve based public key infrastructure to cover the whole user space. Then the sender can select multiple recipients using their public keys with the desired access levels and generate a unique polynomial for that using the Lagrange polynomial interpolation. Next, the content is encrypted using the generated polynomial and multicast it to the recipients. Finally, the recipient can use their private key together with the polynomial to decrypt the received content. Encryption is robust because the elliptic curve cryptography is stronger than the present RSA encryption. Moreover, it is more suitable for mobile devices due to small key sizes. However, the cryptographic libraries have to be improved and optimized in order to make it practical. Further, the applications like email clients, media players, document viewers have to be enhanced to integrate this cryptographic mechanism as an add-on.

Keywords- Elliptic Curve, Multi-key Encryption, Access Control

I. INTRODUCTION

Encryption plays an important role in computer security. With the modern human civilization going back to well over 4000 years, a significant improvement has happened with the use of computing devices for cryptography in the 20th century. With the invention of public key cryptography in 1970's, a set of new dimensions have been added to the field of computer security resulting in significant variation in the way of delivering authentication, integrity verification, key distribution, non-repudiation, digital certificates, etc. With the advent of mobile technology in the recent decade, people meet each other virtually to accomplish their personal and professional tasks. When conducting such virtual meetings over public networks such as mobile and internet, people need to have a robust and secure mechanism that ensures privacy. It has always been a challenge to enforce access control on digital contents especially in a dynamic multi user environment where fixed roles do not exist. However, an encryption method that can open using multi keys separately has not been found yet. In this paper, we present such an encryption method based on existing elliptic curve encryption which will allow users to multicast the same content to multiple users with a single encryption. In this scenario, each recipient has the capability to decrypt the message using his/her private key if permission to access the key has already been granted. Moreover, it will be a robust, fool-proof technology that can be used in devices such as mobile phones that utilizes less computing power.

Another advantage of the technique is its ability to encrypt large contents such as medical images efficiently. (Abdalla, Shavitt & Wool 2000; Armstrong 2006; Ausanka-Crues 2001; Edge & O' Donnell 2016; Godwin 2004; Hazarika 2012; Yao, Lee & Nam 2009; Dahlman 2014).

II. BACKGROUND

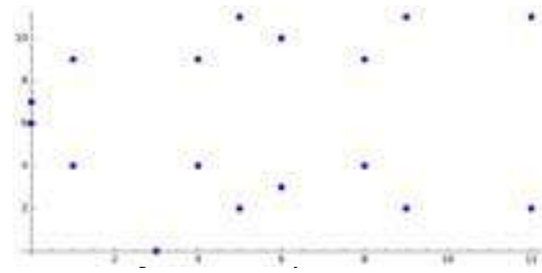
The application of this research is to facilitate medical image encryption in a peer-to-peer real-time mobile collaborative discussion sessions over public accessible mobile networks. Certain mobile apps such as OsiriX, Mobile MIM, etc. allow people to share medical images via public networks. In the recent past, experiments have been carried out to investigate about the privacy and security of sensitive medical data transmitted through unsecured public networks. Moreover, some have used public key cryptography to provide security services such as authentication, integrity verification, non-repudiation, key sharing, etc. Contents were sharing using client server architecture rather than peer-to-peer architecture. Though some of the applications provide role base access control, none of the applications provides multi-key access control mechanisms as present in this research. Further, this research presents the design and implementation of peer-to-peer real-time communication over public mobile networks. (Zhang 2011; Xiaoqin & al 2012; Patel & Bansode 2012).

A. Digital Rights Management

Digital rights management (DRM) is the field where it forces access control policies such as view, save, edit, delete, and distribute digital contents such as eBooks, audio& video songs, films, computer games, cable and satellite televisions, and software CD & DVDs on its users. Though it is not universally accepted, it is necessary to have such a mechanism to protect copy rights as we do with other physical materials. Otherwise it would be difficult for digital content providers to generate revenue from their digital properties as they expect. Thus, digital content providers such as Microsoft, Son, Apple, etc. have tried various techniques over the years and could not be found a strong standard method. Microsoft has tried ten versions of Microsoft Media DRM before 2005 and failed all within a few months. However, the Sony has used an Elliptic curve based DRM to protect SonyPlay Station 3. But it was also cracked within short time due to an issue of its random number generator. Moreover, iTune is enforcing access control on their digital music files. Thus, it limits the number of songs a user can play. (Armstrong 2006; Hazarika 2012; Yao, Lee & Nam 2009; Zhang 2011; Perlman & al 2010).

B. Finite Fields and Elliptic Curves

Numerical computations may contain some round-off errors due to its limitations in floating point representations. However, in cryptography it is not possible to have such errors. As a result, finite fields have been used for most of the cryptographic works. On the other hand, the raising computing power has threatened the existing RSA public key cryptosystems. Therefore, in this research, the novel elliptic curve cryptosystem based on elliptic curve discrete logarithm problem is used. Figure 1 shows the elements of finite field in (Dawahdeh & al 2015; Goo & Lee 2015; Jager & al 2015; Goo & Lee 2015; Joye 2016; Soleymani & al 2013; Boruah & Saikia 2014; Yang & al 2012).



C. Android, Sponge Castel Library and OpenSSL

Android is the most popular mobile platform among the Sri Lankan people today. It is based on Java and Java provides a partially developed cryptographic library for its users. However, the in-built libraries in Java are not sufficient for cryptographic works described in this research. Therefore, a Java based, open source cryptographic library called Sponge Castel has been used to do the cryptographic works within the Android application. Apart from that, a free and open source cryptographic library called OpenSSL has been used to do the other cryptographic works describe in this research. It is mainly used for key generation, signing, verification, and digital certificate management (Bernstein & al 2012; Gozalvez 2015; Zhou & al 2007; Pancholi & Patel 2016; McIntosh 2015).

D. Polynomial Interpolation

Polynomial interpolation is used to find the polynomial that goes through a set of selected points on a space. Though there are several methods to find the interpolated polynomial of a given set of points, the resulting polynomial will not be the same for all the methods (Liu & al 2016; Krishna & al 2016; Perlman & al 2010). Moreover, the Lagrange interpolation takes linear time to find coefficients whereas the Newton interpolation takes quadratic time to do the same. Thus, the Lagrange method for polynomial interpolation is theoretically efficient than the other methods. Moreover, it is a simple algorithm and can be implemented easily. However, in practice, it is not the most efficient algorithm due to its large number of floating point calculations. Besides, researchers have found methods to interpolate a given set of points without passing through specific points if needed. Equation 1 shows the Lagrange interpolation method.

$$p(x) = L_1(x)y_1 + L_2(x)y_2 + \cdots L_N(x)y_N$$

$$L_k(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_{N-1})(x-x_N)}{(x_k-x_1)(x_k-x_2)\ldots(x_k-x_{N-1})(x_k-x_N)}$$

Except;

$(x = x_k)$ *in the numerator and* $(x_k = x_k)$ *in the denominator*

Equation 1. Lagrange polynomial interpolation method

III. METHODOLOGY

First, the architecture of the system was designed. It consists of three main components namely mobile application, servers, and security infrastructure. The mobile application is mainly responsible for applying secure mechanism for access control on digital data and facilitating peer-to-peer video conferencing. Severs are used to store medical images and establish peer-to-peer communications. Finally, the security infrastructure defines the customized public key infrastructure and the proposed access key algorithm.

A. Collaborative Mobile Application

First, by considering the most popular platform of the users, Android has been chosen as the development language of the mobile application. Then a mobile application with collaborative tools was developed to share medical images over the public mobile network. Tools such as real-time video conferencing, interactive whiteboard for real-time collaborative drawings, viewing, zooming, and rotating medical images were developed to facilitate collaborative discussions.

B. File and Application Servers

Medical images were stored in a remote server. L2TP VPN was configured in order to so that the Android Mobile users are allowed to connect to the server over VPN connections and access images securely.

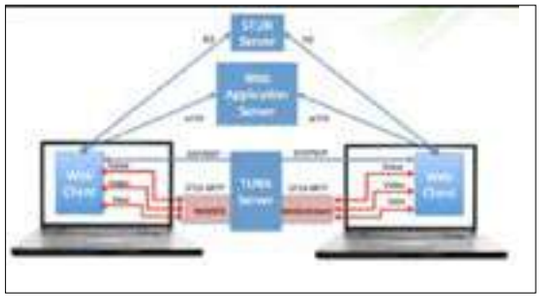


Figure 3. WebRTC server communications
Source: https://bincoder.com/category/webrtc/

Peep-to-peer communications cannot be directly achieved over the public mobile networks due to their local IP addresses. However, if two parties can share their public IP address, then they can communicate with each other directly and securely by opening a VPN tunnel between them. Thus, a third party the WebRTC server has been used to establish the VPN tunnel between peers. Thereafter, peers can do real-time video conferencing and content sharing securely over the public mobile network. Figure 3 shows how the STUN and TURN servers help to establish a real time peer-to-peer communication between peers.

C. Public Key Infrastructure

This method uses the customized public key infrastructure. It is based on the novel elliptic curve cryptography instead of the present RSA encryption. Thus, it is robust and future proof. However, since it is new, not many tools have developed for it yet. Moreover, most of the stakeholders do not have much technical exposure. Therefore, they cannot use command line well. Thus, a Java based GUI tool was developed to manage digital certificates and its associated operations such as generating keys, generating certificate requests, signing certificates and documents, verifying certificates and documents, renewal and etc. Certificate authority hierarchy was established and the certificates were issued to all the stakeholders. However, the libraries are not optimized. Thus, the efficiency of the algorithms cannot be achieved as expected.

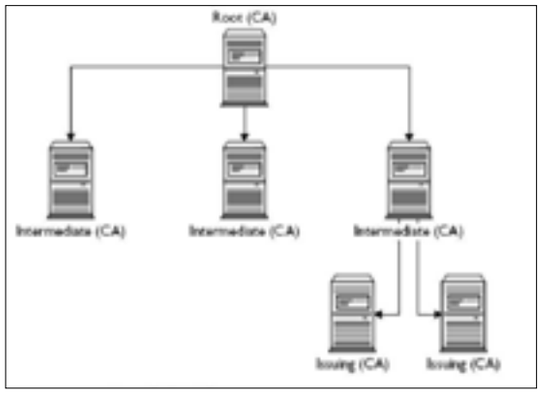


Figure 2. Public key infrastructure
Source: http://www.vce-download.net/study-guide/comptia-networkplus-11.6-remote-access-procedures2.html

D. The Proposed Multi-key Access Control Mechanism

The proposed multi-key encryption algorithm is consisting of four main steps as follows.

1) **Curve Parameters and Modulo Arithmetic:** Once an elliptic curve is chosen, the curve parameters will be extracted and elliptic curve arithmetic will be implemented. Next, the modulo arithmetic operations will be implemented for the chosen number. Moreover, a new user defined data type will be used to handle big integers.

2) **Public Keys and Polynomial Generation:** When a digital content is multicast, it's authorized users and their access levels will be chosen. Their elliptic curve public keys will be extracted from their digital certificates and interpolated polynomial will be generated using the Lagrange interpolation method. $P(x) = ax^3 + bx^2 + cx + d \ mod \ p$ shown in the Equation 2. Figure 4 shows an interpolated polynomial using Lagrange interpolation method.

Equation 2. Generic format of the polynomial

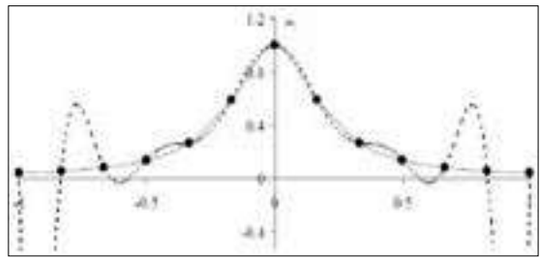


Figure 4. Interpolated polynomial
Source: https://melissacabral.wordpress.com/ 2009/11/29/ lagrange-interpolating-polynomial-2/

3) **Secret Session Key Generation and Encryption:** The generated Lagrange interpolated polynomial will be used to generate a temporary secret session key. This key will be used to encrypt the content that wanted to be multicast using the AES symmetric key algorithm. Thus, it becomes more efficient and robust.

4) **Regenerate the Secret Session Key and Decryption:** Recipients will be used their private key and the received polynomial to regenerate the session key that used to encrypt the content. Thereafter, the obtained AES symmetric is used to decrypt the content and view it the mobile application. Because the content is decrypted with the mobile application, the end users will not be able to access the original content and reuse it as they wish.

IV. RESULTS AND DISCUSSION

A mobile application has been developed for real-time secure video conferencing and medical image sharing. Figure 5 shows a screenshot of the video conferencing app.

Theoretically, a successful multi-key encryption mechanism has been developed. However, due to lack

Figure 5. Video conferencing app

of well-established libraries and standards, the proposed secure mechanism could not be tested practically. For example, a basic computation such as addition should support addition of very large integers, elliptic curve group addition, and modulo arithmetic within a mobile device. However, it is based on the solid mathematical concepts and future proof technologies. These missing libraries will be developed in the future. Thus, there is a high probability of being used it in the near future.

V. CONCLUSION

The necessity of digital rights managements, historical attempts of DRM and their failures have been presented. The challenges of implementing access control in a multi user domain have been identified and emphasized. The proposed secure algorithm will be able to provide multi-key encryption mechanism to enable the access control in a multi user environment. The technical limitations and lack of supporting libraries have been identified and published for the benefit of future researches and the field. Further the end user applications such as email clients, media players, document and image viewers, etc. have to be enhanced in the future to support this secure mechanism as an add-on. However, a theoretical model of a successful, novel, future proof and robust access control mechanism has been invented and presented in this research.

REFERENCES

Abdalla, M, Shavitt, Y & Wool, A 2000, 'Key management for restricted multicast using broadcast encryption', *Networking, IEEE/ACM Transactions on*, vol 8, no. 4, pp. 443-454.

Armstrong, TK 2006, 'Digital rights management and the process of fair use', *Harvard Journal of Law & Technology*, vol 20, p. 49.

Ausanka-Crues, R 2001, 'Methods for access control: Advances and limitations', *Harvey Mudd College*, vol 301.

Bernstein, DJ & al, E 2012, 'High-speed high-security signatures', *Journal of Cryptographic Engineering*, vol 2, no. 2, pp. 77-89.

Boruah, D & Saikia, M 2014, 'Implementation of ElGamal Elliptic Curve Cryptography over prime field using C', *Information Communication and Embedded Systems (ICICES)*, 2014 International Conference on.

Dahlman, EAMGAPSAEA 2014, '5G wireless access: requirements and realization', *IEEE Communications Magazine*, vol 12, no. 52, pp. 42--47.

Dawahdeh, ZE & al, E 2015, 'Modified ElGamal Elliptic Curve Cryptosystem using Hexadecimal Representation', *Indian Journal of Science and Technology*, vol 8, no. 15.

Edge, C & O' Donnell, D 2016, 'Introduction to Cryptography', in *Enterprise Mac Security*, Springer.

Godwin, M 2004, 'What Every Citizen Should Know About DRM', *Public Knowledge New America Foundation, Washington*.

Goo, E-H & Lee, S-D 2015, 'Reconfigurable real number field elliptic curve cryptography to improve the security', *Journal of Computer Virology and Hacking Techniques*, vol 11, no. 3, pp. 123-128.

Gozalvez, J 2015, '5G Tests and Demonstrations [Mobile Radio]', *Vehicular Technology Magazine, IEEE*, vol 10, no. 2, pp. 16--25.

Hazarika, SS 2012, 'Digital Rights Management: A Restrictive Rather than a Defensive Mechanism and the Survival of the 'Fair Use' Doctrine', *Available at SSRN 2180305*.

Jager, T & al, E 2015, 'Practical invalid curve attacks on TLS-ECDH', *Computer Security--ESORICS 2015*, pp. 407-425.

Joye, M 2016, 'Secure ElGamal-Type Cryptosystems Without Message Encoding', *in The New Codebreakers,* Springer.

Krishna, CV & al, E 2016, 'Design Implementation of Composite Field S-Box using AES 256 Algorithm', *International Journal of Emerging Engineering Research and Technology (IJEERT)*, vol 3, no. 12, pp. 43-51.

Liu, X & al, E 2016, 'A Secure Medical Information Management System for Wireless Body Area Networks', *KSII Transactions on Internet & Information Systems*, vol 10, no. 1.

McIntosh, C 2015, 'Cyber-security: Who will provide protection?', *Computer Fraud & Security,* vol 2015, no. 12, pp. 19-20.

Pancholi, VR & Patel, BP 2016, 'Enhancement of Cloud Computing Security with Secure Data Storage using AES', *International Journal for Innovative Research in Science and Technology*, vol 2, no. 9, pp. 18-21.

Patel, JR & Bansode, RS 2012, 'Hybrid Security Algorithms for Data Transmission using AES-DES', *International Journal of Applied Information Systems (IJAIS)*, vol 2, no. 2, pp. 15-21.

Perlman, R & al, E 2010, 'Privacy-preserving DRM', *Proceedings of the 9th Symposium on Identity and Trust on the Internet*, ACM.

Soleymani, A & al, E 2013, 'A Novel Public Key Image Encryption Based on Elliptic Curves over Prime Group Field', *Journal of Image and Graphics,* vol 1, no. 1.

Xiaoqin, L & al, E 2012, 'Application of the Advanced Encryption Standard and DM642 in the image transmission system', *Computer Science Education (ICCSE), 2012 7th International Conference on*.

Yang, CAO & al, E 2012, 'One ECDH key agreement scheme with authentication based on ECDLP', *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, vol 1, no. 24.

Yao, J, Lee, S & Nam, S 2009, 'Privacy preserving DRM solution with content classification and superdistribution', *Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE, IEEE.*

Zhang, X 2011, 'A survey of digital rights management technologies', *last modified: Nov*, vol 28, pp. 1-10.

Zhou, Y & al, E 2007, 'Access control in wireless sensor networks', *Ad Hoc Networks*, vol 5, no. 1, pp. 3-13.

ARTIFICIAL INTELLIGENCE APPROACHES FOR IMPROVED ADAPTABILITY IN AN ADAPTIVE E-LEARNING ENVIRONMENT: A REVIEW

KG Hewa[1] and PPNV Kumara
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]krishnihw@gmail.com

Abstract- The concept of e-Learning, which has emerged with the rapid advancement in technology, is a crucial aspect in the field of education. The major issue with the traditional concept of e-learning is that it delivers information to all students in the same manner, irrespective of their individual learning requirements. Adaptive e-Learning Systems, which emphasise the significance of the differences in individual learning styles in modelling the ideal learning environment, attempts to bridge the gap between the student and the instructor that can be observed in a traditional e-learning environment by identifying and catering to individual learner requirements and capabilities. Artificial Intelligence techniques which have the ability to replicate the decisionmaking process of humans, are significant in the domain of adaptive e- Learning as they can be used to improve the adaptivity of the system. This paper assesses the Artificial techniques; Fuzzy Logic, Neural Networks, Bayesian Networks and Genetic Algorithms, emphasising their contribution towards the concept of adaptivity in the context of Adaptive e-learning. The study indicated an increase in the adaptation of Fuzzy Logic techniques, specifically Type 2 Fuzzy Logic Systems, and Bayesian Networks in the development of the Student Model in order to deal with the uncertainty of learning and student diagnosing processes. The application of Artificial Neural Networks to overcome issues in the existing Adaptive E-learning Systems, has also been identified through this review where the application of feature extraction via the Neural Network approach is an effective methodology to be used in the development of the Adaptation Model of an Adaptive E-Learning System to extract the most appropriate characteristics that can be used to identify learning styles of learners.

Keywords- adaptive e-Learning, Artificial Intelligence techniques, fuzzy logic, Bayesian networks, Neural Networks, Genetic Algorithms

I. INTRODUCTION

The role of information technology in education has changed rapidly and significantly with the introduction of the concept of e-Learning, which provides distance learning opportunities to students all over the world. The focus has been shifted from teacher-centric education to learner centric education where guidance is provided to the learners based on their interests, desires, ability and skills (Lee et al., 2009). In a traditional classroom environment, the instructor can change or adapt his method of delivery according to the learners' behaviour, level of interaction and knowledge level whereas traditional e-Learning provides the same delivery to all students irrespective of their individual learning requirements or preferences (Surjono,2007). Adaptive learning is a new approach to e-Learning, which attempts to bridge this gap which is observed in traditional e-Learning systems by changing the presentation of learning materials to suit the learning style of each individual (Mahajan, R et. Al, 2012). It is based on the assumption that each learner has different learner-characteristics and that they require different educational settings in order to have an effective learning experience (Cronbach and Snow, 1977). An adaptive learning system provides a personalized learning environment taking in to consideration the specific learning requirements and knowledge levels of each individual student.

An Adaptive e-Learning System (AES) consists of three models; Student model, Domain model and an Adaptation model (Cannataro et al., 2002). The student model, also referred to as the learning model or the user model, is the contains all student information such as their domain knowledge, behaviour, knowledge level and other information and is used to adapt the interaction mode of the e-Learning system according to the user's needs (Brusilovsky and Peylo, 2003). The Domain model contains the course content to support adaptive course delivery. It acts as a data repository that consists of topics, contents, pages or nodes and navigation links related to the design structure of the represented data. The adaptive model incorporates the adaptive theory of an AES by combining the domain model with the student model. It defines what can be adapted and how and when it is to be adapted. The levels of abstraction at which adaptation may be defined range from specific programmatic rules to general specifications of logical relationships between Adaptive Learning Environments (Paramythis and Loidl-Reisinger,2004).

The efficiency of the AES relies heavily on the methodology used to collect information regarding the learning requirements and characteristics of students as well as how this information is processed to develop an adaptive and intelligent learning context (Shute and Zapata-Rivera, 2012). AES emphasises the importance of individual differences between learners when attempting to model an ideal e-Learning environment, focusing on identifying and catering their individual learning requirements. Artificial Intelligence concepts have a significant value in the development of an adaptive e-Learning environment as they have the ability to replicate the decision-making process adopted by humans (Frias-Martinez et al., 2004).

This paper assesses the Artificial Intelligence approach to effective learning in an adaptive e-Learning environment. An overview of the artificial intelligence techniques; Fuzzy Logic, Neural Networks, Bayesian Networks and Genetic Algorithms, how they are used in an adaptive learning environment and the manner in which they contribute towards enhancing the adaptability of the system to cater to individual learning styles of the leaners is discussed.

II. RELATED WORK

The concept of adaptation in the context of learning has become a major topic of research in the past few years. Research has shown that the application of adaptation can provide a better learning experience since learners perceive and process information in very different ways. With the advancements in educational technology, specifically the concept of e-Learning, it is important to understand the learning style and preference of learners in order to make learning more effective (Radwan N, 2014). Adaptive e-Learning is a new approach which is based on AI, that can make a traditional e-Learning system more effective by adapting the presentation of information and overall linkage structure to individual users in accordance with their knowledge and behaviour (Esichaikul et al., 2011).

According to Modritscher et al (2004), the adaptation of the teaching and learning process can be divided into four aspects based on a hypothetical e-Learning System as Adaptive Content Aggregation systems, Adaptive Presentation, Adaptive Navigation and Adaptive Collaborative Support. Adaptive Content Aggregation Systems provides students with different content types depending on the teaching and learning styles. Adaptive Presentation refers to the adaptation of the presentation of content within a page by providing prerequisite, additional, or comparative explanations whereas Adaptive Navigation helps the adaptation process by managing personalized views in the content pages. A network-based educational system is used to form a collaborating group of learners in Adaptive Collaborative Support. There are some other studies which explore certain other approaches to adaptive e-Learning. A multidimensional design model which describes specifications needed for an educational environment proposed by Dall'Acqua (2009) examined the conditions that make a learning environment 'adaptive'. Dekson and Suresh (2010) carried out a survey on the various models of adaptive content delivery system and proposed newer methods of delivering adaptive content for an adaptive e-portfolio system. A new approach to integrate learning styles into adaptive e-Learning hypermedia system was presented by Mustafa and Sharif (2011), who assessed the effect of adapting educational materials individualized to the student's learning style.

The aim of an adaptive e-Learning environment is to tailor the overall learning approach in order to fulfil the learning

requirements of students (Essalmi et al., 2010). An adaptive system adjusts itself to suit particular characteristics and requirements of the learner. Artificial Intelligence (AI) techniques are significant in the domain of adaptive e-Learning as they have the ability to develop and replicate the decision-making process of humans. The application of AI within e-learning has the potential of creating a realistic and interactive educational environment. AI approaches are used in adaptive educational concepts in a variety of ways. For instance, in some systems, the main focus is to examine and assess student characteristics in order to generate student profiles with the intention of evaluating their overall level of knowledge to be used as a basis in the Student Modelling process (Gamboa 2001, Gertner and VanLehn 2000). AI approaches are also used to facilitate the diagnostic process completion so that course content can be adjusted to cater to the learning requirements of every student, and some of them are used to learn from student behaviours to adjust the prescribed software pedagogy (Xu et al. 2002, Moreno et al. 2005).

III. ARTIFICIAL INTELLEGENCE TECHNIQUES FOR ADAPTIVE E-LEARNING

Adaptive educational systems emphasise the significance of individual learning requirements of students in the process of E-learning where it seeks to bridge the gap between the student and the instructor in the traditional e-learning environment. Artificial Intelligence approaches are regarded as valuable tools in many contexts as they have the ability to develop and replicate the decision-making process of humans. The AI techniques; Fuzzy Logic, Neural Networks, Bayesian Networks and Genetic Algorithms and their deployment in the adaptive learning context are discussed in this section.

A. Fuzzy Logic

Initially presented by Zadeh (1965), Fuzzy Logic is considered an efficient user modelling technique in adaptive e-Learning, specifically as it can mimic the logical reasoning process of humans (Bih, 2006). Learning-teaching behaviour is represented in a human readable and linguistically interpretable manner through the use of fuzzy rules. The Fuzzy Logic System (FLS), as represented in Figure 1, consists of for components: Fuzzifier, Rule base, Inference engine and Defuzzifier (Mendel, 1995).

Figure 1. Overview of a Fuzzy Logic System (Almohommodi and Hagras, 2013)

A framework geared towards user modelling, based on FLS, can be developed in the context of adaptive e-Learning, inducing simplified reasoning for users as well as developers (Jameson 1996, Kavi et al. 2003). FLSs can be adopted in order to evaluate and assess learning and knowledge-related outcomes, more precisely examine task objectives, multiple criteria assessments etc. FLS can also be deployed in the Student Modelling process in the development of AES, so that the course content can be adopted in order to cater to individual learning styles. Type 2 FLSs is a new classification of fuzzy systems, used to convey numeric and linguistic uncertainty and can be proposed to directly model and reduce the effects of uncertainties. Extended form type 1 FL, type 2 FL provides a methodology for handling various sources of numeric and linguistic uncertainty which prevail in e-Learning environments.

Xu et al. (2002) presented a profiling system which employs the multi agent approach where fuzzy models were created for content and students based on a dynamic plan formally predefined for a specific individual. This framework was developed through profile abstraction and the content framework which was devised and created through the use of fuzzy links between the subject areas and the knowledge of individuals which were established to be utilized in order to determine learning adaptation. Students are able to receive personalized learning material, quizzes and advices through the proposed system. Although results based on a questionnaire regarding the personalization function of the proposed system as opposed to the traditional learning environment indicate that the proposed system makes a great improvement on personalization of learning, the awareness of the system of the changes in the learning styles and other traits of the student has to be taken into consideration in order to maintain the personalization levels of the system.

Gupta and Dhawan (2012) formulated a unique mathematical model for academic performance synthesis and analysis of an institute. FL is applied in the model for crafting decisions where decision vectors are deployed to deliver suggestions on improving the academic performance of the institute. An FL based student model, which removes the arbitrary specification of precise numbers and facilitates modelling at a higher level of abstraction, was developed by Goel et al. (2012). The proposed student model is designed based on Knowledge tracing and Fuzzy inference where two-rule based systems; one for student knowledge diagnosis and the other for the prediction of a parameter for student performance, have also been designed. However, the results demonstrate overfitting and lack of precision initially, where the integration of machine learning techniques will help in improving the accuracy of the model at the initial stage. Research proposed by Almohommadi and Hafras (2013) makes use of a self-learning mechanism that generates a FL-based model from the available data. The practical and theoretical environment proposed, incorporated a system which determined student engagement levels based on visual information utilization, which differed from the traditional method of using sensors. Type 2 FL models was used to improve the knowledge delivery to students based on their individual characteristics. Continuous adaptation was maintained in the proposed system in order to ensure that the generated models adapted to the individual learning styles, preferences and needs of the students. Although experiments were carried out targeting only 17 post-graduate students, the system facilitated knowledge delivery improvements to individuals, thus helping to create better performance than type 1 fuzzy systems.

B. Neural Networks

Information processing and problem solving can be performed through the use of a large number of interwoven neurons or components, known as an Artificial Neural Network (ANN). They are specifically in modelling human behaviour due to their ability to process and produce complex information (Frias-Martinez et al., 2004). ANNs which are considered as an effective mechanism to solve various problems, particularly in the field of science, also contribute to the educational field specifically in the context of adaptive e-Learning.

Research has been carried out with regard to the use of ANNs in the field of education, specifically in the context of adaptive e-Learning. Villaverde et al. (2006) described an approach based on a three-layered feed-forward ANN to recognize the learning styles of students based on the actions they have performed in the e-learning system. The proposed automatic style recognition mechanism facilitates the gathering of information about learning preferences where the proposed algorithm uses the recent system usage history to recognize changes in learning styles. This approach has the ability to improve adaptive e-learning environments by identifying individual learning styles of students and presenting them with a customized learning environment to meet their individual learning requirements. A personalized multi-agent e-learning system based on item response theory and ANN which presents adaptive tests and personalized recommendations, was proposed by Baylari and Montazer (2009). The proposed system contains four agents; activity agent, planning agent, test agent and remediation agent. The Remediation agent, designed using ANN, specifically back-propagation, acts as a human instructor with the ability to diagnose the learning problems of the learner and recommend necessary learning materials to the learner. It also has the ability to guide the learner through a remediation session in order to improve his/her learning problems. The proposed system is also capable of determining the learners' ability based on explicit responses on tests to present him/her with a personalized and adaptive test based on that ability. Although the experimental results demonstrated an 83.3% precision in providing personalized and appropriate course material recommendations, this result was obtained through the use of only 30 test cases, thus the level of compatibility between the level of personalization provided by the system and the actual learning style of the learner cannot be clearly established.

Chaplot et al (2016) presented a new adaptive learning system architecture based on ANN which has the ability to handle multi-concept items for effective prediction of student performance and select practice items of optimal difficulty personalized to the student's skill level. The proposed system overcomes the two shortcomings of existing AES; inability of the Student Model to handle multi-concept problems and the inability of the Domain Model to systematically select problems of appropriate difficulty for the student to maximize learning gain, through the use of ANN.

PROCEEDINGS

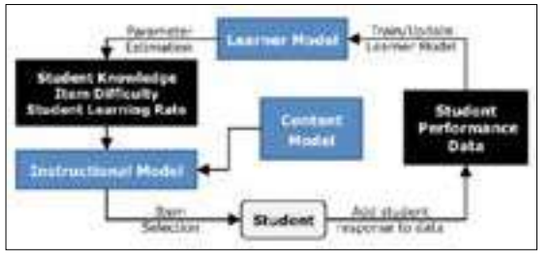


Figure 2. Proposed Adaptive Learning System Architecture (Chaplot et al., 2016)

In order to handle multi-concept problems, the Student Model or the Learner Model in the proposed system architecture (Figure 2), is designed using an ANN, which does not assume any relationship between inputs, contrary to previous methods. Hmedna et al. (2017) describes a methodology on how ANN can be used to identify the learning styles of learners based on the actions performed by them in Massive Open Online Courses (MOOC). The proposed system has a six-stage approach; Data collection, Pre-processing, Feature extraction, Classification, Learner profile and Adaptation. The feature extraction method which is applied after pre-processing extracts the most appropriate characteristics that can be used to identify learning styles of learners. A Neural Network is used in the classification phase (Supervised classification) to detect and recognize learning styles. This methodology can be adapted by AES in identifying individual learning styles thereby helping to improve the adaptability of the system. Supervised classification approach used in this project is a significant step where the learning material can be provided to learners in a more effective manner according to their individual learning style, but the full advantage of this approach cannot be achieved due to the fact that this system only has the feature of recommending content based on the learning style of the learner.

C. Bayesian Networks

A Bayesian Network(BN) is a Direct Acyclic Graph (DAG) explaining the probability distribution in a manner which allows proficient probability dissemination as well as an accurate representation (Gamboa 2001, Moreno and Moreno 2005). It is widely used for student modelling in intelligent learning systems.

An e-Learning system with an integrated learner assessment module that implements personalized content delivery through the use of case-based reasoning and BN was developed by Mangalwede and Rao (2010). A set of questions for a particular course are considered where the set of questions are divided in to three categories as easy, moderate and difficult. The questions are structured in an acyclic graph where the probabilistic relationship between two successive questions are captured using BN. This approach attempts to test the real understanding and the capability of the learner where the motivation is to never ask a question that the learner can answer effortlessly. The learner's performance in the given assessments will be used in order for the process of generating and delivering appropriate learning content in the next session.

Jeo and Su (2011) presents an approach which introduces a new rule specification language and provide a user interface for the domain expert to specify the condition part of an adaptation rule probabilistically. A Bayesian model is used in this approach to enable the evaluation and apply proper adaptation rules to tailor an instruction for each new learner in the presence of data anomalies. The accumulation of 'group data' in the proposed approach will improve the accuracy of evaluating the next new learner enabling continuous improvement of the adaptive capability of the system. However, the proposed methodology will be much beneficial if the structural learning approach, acquiring the structure of a Bayesian Model based on learners' data, was also explored.

Bachari et al (2014) proposed a new personalized education system based on the personality of the learner. The proposed framework consists of three elements, namely the domain model, learner model which includes a preference engine and the pedagogical model which includes an adaptive engine model and a revised strategy model. The adaptive strategy engine sets a learner in one of four kinds of independent groups (Figure 3), which are derived based on the preference engine which detects and stores preferences in student model according to Myers-Briggs Type Indicators (MBTI) Tools.

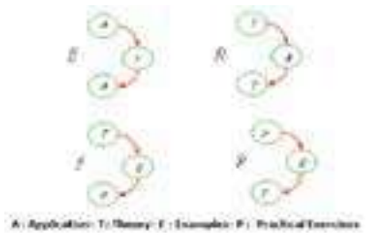


Figure 3. Adaptive Taxonomy: LS dimensions and TS relationships (Bachari et al, 2014)

The revisited strategy engine helps determine the suitability of a particular teaching style to a given learning style using Dynamic Bayesian Network Classifiers. The system offers a suitable teaching style and if the learners grade is less than 60%, the system presents another teaching style.

D. Genetic Algorithms

Genetic algorithms(GA) are optimization algorithms based on the theory of the evolution of the species by Charles Darwin, where the purpose is to seek an extreme of a function defined on a space of data (Azough et al., 2010). According to Davis (1991) the optimization process is performed by GA in four stages: initialization. Selection, crossover and mutation. GAs have become increasingly popular in the development of educational systems as they are especially helpful in understanding the preferences and requirements of the end user (Drigas et al., 2009).

A personalized approach for curriculum generation supported by a GA-based module to facilitate a personalized generation of learning paths, was proposed by Huang et al. (2007). The learning path generation proposed, has the ability to consider the curriculum difficulty level and the curriculum continuity of successive curriculums simultaneously, while implementing a personalized curriculum generation in learning processes. The GA based module of the proposed system, composed of a generation engine and an XML-based knowledge description, allows the generation of the personalized learning path for web-based instruction. This system provides a significant contribution towards the involvement of GA's in the adaptive eLearning process where it has the ability to provide personalised material and quizzes as well as generate curriculums accordingly. A drawback of the proposed system is that it does not accept user feedback which is an important feature which enhances the adaptive nature of a learning environment.

A GA based adaptive learning scheme for context aware eLearning was proposed by Bhaskar et al (2010). Three levels of contexts of learner; Content level, Presentation level and Media level context, are considered when generating a learning scheme in the developed system, where the Content layer deals with learning path generation, the presentation level with learner preferences and intentions and the media level with the media preferences of the learner. The system also includes four context tracking modules and a genetic based adaptive learning scheme generation algorithm module where the learning path generation algorithm is enhanced and evolved to a leaning scheme generation which generates a learning path accommodating the entire learner's context. This learning scheme generation algorithm is designed to be genetic as the various learner's context parameter values are viewed as constraints to be fulfilled in the learning scheme generation. When compared to the system proposed by Huang et al. (2007) this system has less features considering that the former has the capability of providing suggestions to learners.

Azough et al. (2010) developed an adaptive educational system based on the modulization of the description of learning resources which has the ability to provide the path that is most adapted to the learner profile by using optimization algorithms. The main focus of the system is to establish an optimal path from the learner profile extending to the learning objective based on intermediate courses through the utilization of GAs. It helps a learner to be more autonomous, to have a better comprehension of the course and to better apprehend and manage his learning process. This system has the groundwork to provide personalized learning content as well as quizzes and suggestions by further evaluating and enhancing the mechanism that has been used in order to generate pedagogical paths.

An AES which allows the learner to take courses adapted to his profile and to the pedagogical objectives set by the instructor was described by Madani et al. (2017). A GA, which consists of iterating the three operations; reproduction, crossing and mutation, have been used for the adaptation process to provide the learner the concepts to be learned in an optimal manner by seeking the objectives most adapted to his profile. A measure called a period of activity is introduced to calculate the period of day in which the learner was active where all the publications that are done during this period is filtered in terms of teaching and education. A classification of the collected publications is done where they are classified in to three as motivate, demotivate or neutral. A parallelization of the GA in order to find the optimal pedagogical objective for the learner as well as a parallelization of the classification of publications for finding the sentiment of the learner during the period of activity is also performed. The concept of parallelization as well as sentiment analysis which have been used in this system are new approaches that have not been used in other existing AES and are of much significance.

IV. CONCLUSION

The concept of adaptive e-learning emphasises that the learning process of one learner differs from that of another. The establishment of accurate student models and profiles by modelling the personality traits, skills and the knowledge level of students is crucial when developing an adaptive learning environment. The AI approaches; Fuzzy Logic, Neural Networks, Bayesian Networks and Genetic Algorithms, were assessed in this paper in relation to the concept of adaptive e-learning. This review shows an increase in the adaptation of Fuzzy Logic, specifically Type 2, and the application of Artificial Neural Networks to overcome the major issues in the existing AES systems. The use of Bayesian Networks in the development of a student model in order to deal with the uncertainty of learning and student diagnosing processes has also been identified through this review where the use of Genetic Algorithms in the adaptation process to provide the learner the concepts to be learned in an optimal manner has also been assessed. From the AI methodologies discussed, feature extraction via the Neural Network approach can be considered as an effective methodology to be used in the development of the Adaptation Model of an Adaptive E-Learning System where the most appropriate characteristics that can be used to identify learning styles of learners can be extracted. This methodology provides an accurate identification of a learners individual learning style which is essential in an AES in order to deliver course materials to each student in an effective manner. This study can be further carried out by way of implementation where an adaptive eLearning system can be developed with the use of ANN, as it was identified as the more effective methodology to be used, in order to determine the effectiveness of the above mentioned methodology.

REFERENCES

Almohommadi K, Hagras H (2013), An Interval Type-2 Fuzzy Logic Based System for Customized Knowledge Delivery within Pervasive E-Learning Platforms, *Proceedings of the 2013 IEEE International Conference on Systems, Man and Cybernetics*, 2872-2879

Almohommadi K, Hagras H (2013), An adaptive fuzzy logic based system for improved knowledge delivery within intelligent E-Learning Platforms, *Proceedings of the 2013 IEEE International Conference Fuzzy Systems(FUZZ-IEEE)*, 1-8

Azough S, Bellafkih M, Bouyakhf EI (2010), Adaptive E-learning using Genetic Algorithms, *IJCSNS International Journal of Computer Science and Network Security*, 10(7):237-244

Bachari EE, Abdelwahed EH, Adnani ME (2014), Design of an Adaptive e-Learning Model based on learner's personality, *Ubiquitous Computing and Communication Journal*

Baylari A, Montazer GA (2009), Design a personalized E-Learning system based on item response theory and artificial neural network approach, *Expert Systems with Appications*,36(4):8013-8021

Bhaskar M, Das MM, Chithralekha T, Sivasatya S (2010), Genetic Algorithm Based Adaptive Larning Scheme Generation for Context Aware E-Learning, *International Journal on Computer Science and Engineering*, 2(4):1271-1279

Bih J (2006), Paradigm shift-an introduction to fuzzy logic, *IEEE Potentials*, 25(1):6-21

Brusilovsky P, Peylo C (2003), Adaptive and Intelligent web-based educational systems, *International Journal of Artificial Intelligence in Education*, 13: 156-169

Chaplot DS, Rhim E, Kim J (2016), Personalized Adaptive Learning using Neural Networks, *Proceedings of the Third Annual ACM Conference on Learning @ Scale*, 165-168

Cannataro M, Cuzzocrea A, Mastroianni C, et al. (2002), Modeling adaptive hypermedia with an object-oriented approach and XML, *Proceedings of 2nd International Workshop on Web Dynamics*, Honolulu, Hawaii

Cronbach LJ, Snow RE (1977), Aptitudes and Instructional Methods: A handbook for research interactions, New York: Irvington

Dall'Acqua L (2009), A model for an adaptive e-Learning environment, *Proceedings of the World Congress on Engineering and Computer Science(WCEES)*

Davis L (1991), Handbook of Genetic Algorithms, Amsterdam: Van Nostrand Reinhold

Dekson, DE, Suresh ESM (2010), Adaptive E-Learning techniques in the development of teaching Electronic Portfolio – *A survey, International Journal of Engineering Science and Technology*, 2(9): 4175-4181

Drigas AS, Argyri K, Vrettaros J (2009), Decade Revie, Artificial Intelligence Society, Knowledge, Learning, Development and Technology for AII, 49:552-564

Dunham MH (2006), Data Mining: Introductory and advance topic, Pearson Education, India

Esichaikul V, Lamnoi S, Bechter C (2011), Student Modelling Adaptive E-Learning Systems, *Knowledge Management & E-Learning: An International Journal*, 3(3): 342-355

Essalmi F, Ayed LJB, Jemni M, et al. (2010), A fully personalization strategy of E-learning scenarios, *Computers in Human Behaviour*, 26(4): 581-591

Friaz-Martinez E, Magoulas G, Chen S, Macredie R (2004), Recent Soft Computing Approaches to User Modelling in Adaptive Hypermedia, *Proceedings of International Conference on Adaptive Hypermedia and Adaptive we based systems*, 104-114

Gamboa H (2001), Designing Intelligent Tutoring Systems: A Bayesian Approach, *Proceedings of Ana Fred 3rd International Conference on Enterprise Information Systems (ICEIS 2001)*, 451-458

Goel G, Lalle S, Luengo V (2012), Fuzzy Logic Representation for Student Modelling: Case study on Geometry, *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, 428-433

Huang MJ, Huang HS, Chen MY (2007), Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach, *Expert Systems with Applications*, 33(3):551-564

Hmedna B, Mezouary AE, Baz O, et al. (2017), Identifying and tracking learning styles in MOOCS: A neural networks approach, *International Journal of Innovation and Applied Studies*, 19(2):267-275

Jameson A (1996), Numerical uncertainty management in user and student modelling: An overview of systems and issues, *Use Modelling and User-adapted Interaction*,5(4):103-251

Joen SS, Su SYW (2011), Deriving Prior Distributions for Bayesian Models Used to Achieve Adaptive E-Learning, *Knowledge Management and E-Learning: An International Journal*, 3(2):251-270

Kavi A, Pedraza-Jimnez R, Molina-Bulla H, et al. (2003), Student Modelling Based on Fuzzy Interface Mechanisms, *Proceedings of IEEE Region 8 EUROCON 2003, Computer as a Tool*, Ljubljana, Slovenia

Lee MW, Chen SY, Chtysostomou K, et al. (2009), Mining student behaviour in web based learning programs, *Expert System Applications*, 36(2):3459-2464

Madani Y, BengourramJ, Erritali M, et al. (2017), Adaptive e-learning using Genetic Algorithm and Sentiments Analysis in a Big Data System, *International Journal of Advanced Computer Science and Applications*, 8(8):394-403

Mahajan R, Sodhi JS, Mahajan V 2012, 'Mining User Access Patterns Efficiently for Adaptive e- Learning Environment', *International Journal of eEducation, e-Business, e-Management and eLearning*, pp. 277-279

Mangalwede SR, Rao DH (2010), Application of Bayesian Networks for Learner Assessment in E-Learning Systems, *International Journal of Computer Applications*, 4(4): 23-27

Mendel J (1995), Fuzzy logic system for engineering: A tutorial, *Proceedings of IEEE*, 83(3):345-374

Modritscher F, Barrios VMG, Gutl C (2004), Enhancement of SCORM to support adaptive E-Learning within the Scope of the Research Project AdeLE, *Proceedings of World Conference on E-Learning*

Moreno F, Moreno M (2005), Using Bayesian Networks in the Global Adaptive E-learning Process, *EUNIS 2005*, Manchester,1-4

Mustafa YEA, Sharif SM (2011), An approach to adaptive E-Learning hypermedia system based on learning styles (AEHS-LS): implementation and evaluation, *International Journal of Library and Information Science*, 3(1):15-28

Paramythis A, Loidl-Reisinger S (2004), Adaptive Learning Environments and e-Learning Standards, *Electronic Journal of e-Learning*, 2(1): 181-194

Pires JM, Cota MP (2012), Evolutive Platform-A Genetic E-Learning Environment

Radwan N (2014), An Adaptive Learning Management System Based on Learner's Learning Style, *International Arab Journal of e-Technology*, 3(4):228-234

Shute VJ, Zapata-Rivera D (2012), Adaptive educational systems, in P.Durlach(Ed.), Adaptive Technologies for Training and Education (pp 7-27), New York: Cambridge University Presity

Surjono HD (2007), 'The design and implementation of an adaptive e-learning system', *The International Symposium Open, Distance, and Elearning (ISODEL)*, Denpasar, Indonesia

Surjono HD (2014), The Evaluation of a Moodle Based Adaptive e-Learning System, *International Journal of Information and Technology*, 4(1): 89-92

Villaverde JE, Godoy D, Amandi A (2006), Learning styles' recognition in E-Learning environments with feed-forward neural networks, *Journal of Computer Assisted Learning*, 22:197-206

Quinlan JR (1986), Introduction of Decision Tree, *Journal of Machine learning*, 1:81-106

Xu D, Wang H, Su K (2002), Intelligent Student Profiling in Fuzzy Models, *Proceedings of the 35th Hawaii International Conference on System Science(HICSS)*, Hawaii, USA.

FOSTERING SOCIAL ENGINEERING AWARENESS: PROACTIVE MODEL

HAHV Halwatura[1], WGCI Priyadarshana[2], and T Samarasinghe[3]
[1]Junior Information Security Analyst, DPIT,Battaramulla, Sri Lanka
[2]Network Engineer Trainee, People Bank Innovation center, Colombo 01, Sri Lanka
[3]Associate Engineer, Lanka Bell, Sri Lanka
[1]halwatura.vihanga@gmail.com

Abstract- Social Engineering aims to trick users into revealing sensitive information by making use of their lack of literacy in Social engineering tricks and the limited or no technical mechanisms on their systems to protect against such attacks. The motive of this research is to check the awareness and perceptions on social media of employees from the Information technology sector as well as the other sectors in an equal proportion. This paper shows a series of results which shows the weak points of defending against Social engineering attacks as an individual and in an organizational point of view. The methodology used to conduct this research was an online survey which was sent through email and social media and was successfully completed by 118 people and rejected by approximately 50 people. The awareness or the need of training to identify Social engineering tricks can be clearly seen by the analysed results. As a solution to this escalating issue, this paper suggests a model which is named as 'Proactive model A' that can be used by individuals as well as organizations to mitigate the risks of Social Engineering attacks by implementing the model in their policies and training programs so that it can help in minimizing the damage to critical assets of the organizations.

Keywords- Social Engineering, proactive defence, Cybersecurity

I. INTRODUCTION

People are the weakest link in technology, they might not be a device or a machine but they are the users and although the technology is evolving faster than we think today there is a human involvement in all its processes. (Streeter, n.d.) This is the main reason why Social engineering is an involving threat to all organizations regardless of how small or big the organization is. Social engineering is a huge threat in the modern world but do you think that it is a modern way of attacking in the Cyberspace? If we think up closely it is similar to the situation where the Greeks gifted a horse to the Trojans as a peace symbol but the wooden horse had warriors hidden to destroy their city. ("Social Engineering," 2018). The social engineering scams today has the similar story but using different techniques. According to an organization the Logical Front, 62% of the business have faced phishing and Social engineering attacks and according to their statistics it was the second in the list of attacks. ("6 Security Threats to Look Out for in 2018," 2017). No matter how hard the professionals in Cybersecurity come up with security mechanisms, if the users are vulnerable it can damage the critical assets of an organization or your personal data. Even if the computer systems or the network is patched, updated and secured in all ways, targeting the people in the relevant area can breach all the security mechanisms of the organizations. (Mouton et al., 2015)

Social engineering attacks are not only about the weaknesses of humans that has led to its increase but also the factors such has it needs low technology involvement, low cost and its simplicity in carrying out attacks. The main types of social engineering techniques that most that most people are aware about is email phishing which involves in sending emails which looks authentic but aims to steal users' sensitive information or digital secrets. (Suganya, 2016) In today's world, that's not all as there are other ways to carry out the same task such as by voice calls, social media as your friends, pretexting and many more. (Banu and Banu, 2013)Another thing to highlight is that Social engineering is not restricted to external attacks, it can also happen as an internal attack. In fact, an article in Digital Guardian says that 63% of these attacks come from internal sources ether from errors, control or frauds. ("Social Engineering Attacks," 2015).

This paper focus on proving the weakness and the lack of awareness and knowledge some people have on Social engineering, all data are collected through a survey and analysed. The final recommendations are depicted through the Proactive model A.

II. RELATED WORK AND FINDINGS

To conduct a social engineering attack, one need not have thorough technical knowledge, it is a process of handling with the human psychology which contains many emotions from shyness to curiosity. Knowing how to deal with these emotions might allow the attacker to get information valuable to them by the users. Sometimes getting this information itself might not be the attack but just an insight to how the major attack must be done. (Cert UK, 2015) Recent social engineering attacks include spear phishing which is involved in directly targeting a specific user whereas the other methods known as the water holing is used to attach a malware to a website where the employees of an organization is expected to get infected. (Krombholz et al., 2015) Four of the common tactics used by the Social engineers to steal information is to attract the users by being confident in what they do, next they try to build your trust by offering something, and then they might make some humor and finally they will request what they want along with a legitimate reason. (Luco, 2013) Despite all these facts, the ways to reduce social engineering impact to our data are limited. In a study it says that although the employees have taken trainings, they still share passwords so humans are very vulnerable. (Cazier and Botelho, 2007) The attack mechanisms are very advanced today, although .exe files cannot be sent through mails due to security reasons, they can still be sent by a zipped folder which can result in a successful attack. (n.d.)

The best way to keep away from social engineering tricks is to be aware and conduct trainings for the employees in the company and make sure that the trained matters are all executed in the practical real world scenarios. (Smith et al., 2009).

III. METHODOLOGY

To check the level of literacy among the people from different fields, this research was initiated. The data collection method used to conduct this survey was a Google survey which was distributed to people from different categories. The survey included quantitative and qualitative questions. The distribution of the survey was done via Email and Social engineering sites such as Facebook, Viber and WhatsApp. The survey was made as simply as possible for the respondents to give answers quickly and easily. The first few questions were related to biography and following were the technical questions. A total of 20 questions were there on the survey. This survey was conducted for five months from December 2017 to May 2018 and it has responses from a 118 people and approximately 50 responses have been rejected due to the fact that the respondents weren't aware about the topic. The questions were made to capture the social engineering behaviour in their personal information and also in their corporate environment.

IV. RESULTS AND DISCUSSION

Figure 1. Field of work/study

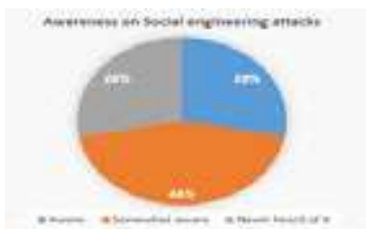


Figure 3. Awareness on Social engineering attacks

All analysed results are presented in this section with a description and a figure containing the statistics.

Figure 6. Awareness if your information is used for social

A majority of the respondents were from the age category 18-36 years of age which was 96% from the total statistics where as 2% was from 37 to 48 category and another 2% was from the 49 to 67 category. Their field of career was divided as 6% was from the Cybersecurity field, a majority was from the IT field with a percentage of 49 and the second highest was the Other field or the fields that relates to non IT category which consisted of 45% as depicted in the Figure 1. Majority of the respondents of these fields are represented by undergraduates with a percentage of 44. And in second are the front line professionals and the executives.

Figure 2 shows the percentages of respondents who knows how to identify an email scam and who doesn't. 61% has answered that they know how to identify scams whereas 39% does not know how to.

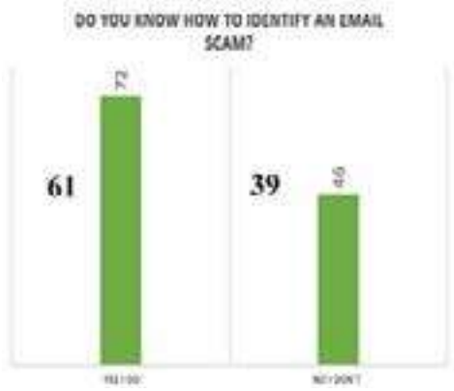


Figure 2. Positivity on identifying an email scam
Source: Author

According to the respondents the overall awareness of Social Engineering attacks are depicted in Figure 3. 28% are aware about social media attacks, 44% are somewhat aware whereas another 28% has never even heard of it.

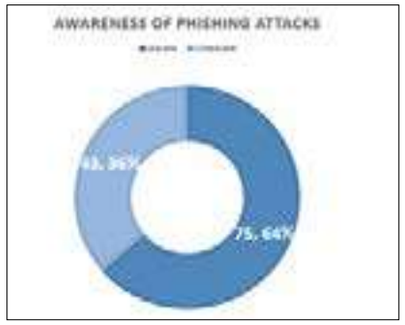


Figure 4. Awareness of phishing

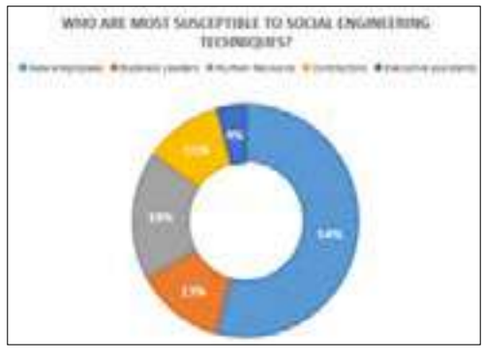


Figure 8. who are most susceptible to social engineering?

The awareness of the most famous social engineering technique phishing attacks can be analysed as the following figure 4. There were 36% who were unaware and only 64% of the respondents were aware about this popular attack.

Figure 5 shows the results for the questions whether the organization they are working for has ever been targeted by a social Engineering attack and a majority answered that they do not know with a percentage of 65%, and 16% answered yes with confidence whereas only 19% answered No.

Another set of answers in Figure 6 showed that only 27% will know if their details are used in a Social engineering attack and 23% said they will not know if their information is used by someone whereas a majority of the people with a percentage of 50% says that they will have no idea if their information is taken for social engineering by a hacker.

Since social engineering is not only about the cyber world so a question regarding social engineering was given on whether they have a method to validate their bank when they get a call from them and as shown in Table 1 a majority replied No with a percentage of 55.9% and No with a percentage of 44.1%.

Table 1. validating your bank/utility supplier

Do you have a method to avalidate your Bank or Utility supplier when they call?	
Yes	48.3%
No	51.7%

A question on how trustful are the employees in their organization in a social engineering attack showed that a majority of the respondents with a 61% think that the some employees might disclose sensitive information where as 16% some employees might definitely disclose information and only 23% thinks that the employees will not give put information as shown in Figure 7.

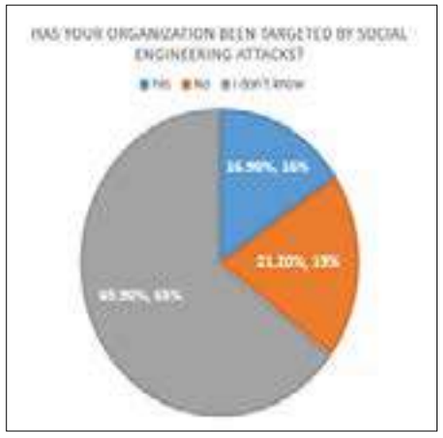


Figure 5. Organizational impact on social engineering

According to the respondents, figure 8 shows the most susceptible people in an organization to Social engineering are the new employees with a percentage of 54% of the respondents, next are the human resource personnel agreed by 18% of the personnel, third highest are the business leaders with a 13% of the responses.

As shown in Table 2. According to the respondents in a case of cyber security incidents, a majority of the people with a percentage of 55.9% doesn't know whom to contact for help, only 44.1% knows the relevant authorities.

Table 2. who do they ask for help?

What is your opinion on the most common source of social engineering threats?	
Email	26.2%
Social Networking Sites	55.1%
Insecure mobile devices	10.2%
Other	8.5%

Source: Author

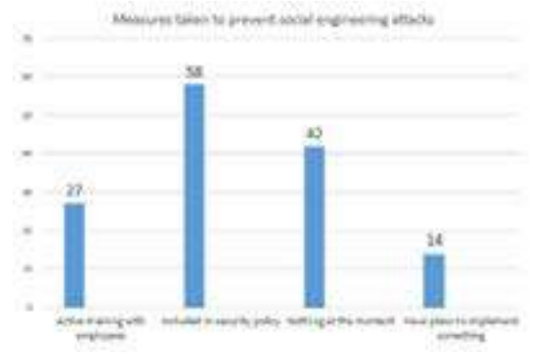


Figure 9. Measures taken to prevent social engineering attacks
Source: Author

Cybersecurity does not depend on the organization since at some point they get involved with the cyberspace whether it's online or offline, the employees must know how to defend themselves against any attack. Therefore, a question on the mechanism they have gone through as an employee to be protected in cyberattacks as such was asked as shown in Figure 9 and 84% respondents replied positively saying they have some security mechanisms whereas the rest of the 56% replied saying either they have plans to implement them or nothing at the moments which is not very positive in this new era of technology.

The survey consisted of a few open ended questions and one was the definition of social engineering in their own terms. Since this was a questions which cannot be answered by all, it was left as a not required questions so only 41 people answered and apart from 3 responses all others were correct but this does not even count 50% of

the responses therefore it shows a weakness in literacy of this term. Following this question was to list three common social engineering attacks and 40 respondents answered to this with the common attack being phishing but 9 responses had to be disregarded due to the fact that they did not have an idea about the question. Having open ended questions in the survey helped to get a better overview on the understanding of the area and scope of knowledge they have.

Table 3. who do they ask for help?

What is your opinion on the most common source of social engineering threats?	
Email	26.2%
Social Networking Sites	55.1%
Insecure mobile devices	10.2%
Other	8.5%

Source: Author

As stated by the respondents in Table 3 the most common source of social engineering attacks are social media websites and emails, insecure mobile devices come third along with other options as the fourth. This might be right because the use of social media is high nowadays but when it comes to a work environment, where social media is not a main source of communication, the main source can be emails and other options such as internal attacks and many more therefore they cannot taken as granted.

The respondents were asked about the way how they can identify a phishing link or email as shown in Figure 10.

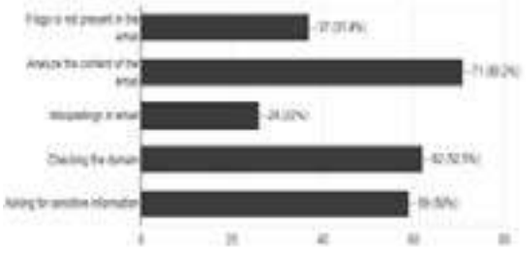


Figure 10. Identifying a phishing email/link
Source: Author

The answers included; If the logo is not present in the email which has percentage of 31.4% and the majority with 60.2% answered Analysing the content of the mail, 22% answered Misspellings in the email, 52.5% answered Checking the domain and 50% answered asking for sensitive information. The answers were different and the percentage of the results varied but this question was a trick to check the literacy rate of how they will react to such kind of email but in reality all the options given in this question but be considered when determining the authenticity of the message or link. The results were taken by both information technology personnel as well as professionals from other fields therefore the results are more reliable.

This study shows an effective message regarding Social engineering and the literacy rate along with their perception of Social engineering and by analysing these results we can come to an obvious fact that the awareness on Social engineering is very low. Approximately 50 responses were rejected and the reason for that was that the technical word Social engineering was not a familiar word and when they opened the link, they couldn't understand the questions because they are less literate about security in technology.

The best way to mitigate the risks in Social engineering is to train or conduct awareness programs to educate the employees and the general public about these tricks that scan happen so that they will be more careful in the future. Majority of the respondents have not been hacked and at the same time majority of the people are unware or somewhat aware about the malware, only 28% are fully aware about it. Therefore, it only means that there's more space for them to get hacked quickly. When it comes to an organizational point of view, 35.6% of the respondents say that no mechanism is used to control it at the moment and 12% said they have plans and the rest have security mechanisms such as policies in place to prevent such attacks along with active training. The new employees in a company is more likely to get targeted in social engineering attacks as shown in the results positively by the respondents, this is because they might not be familiar with the work environment and the appropriate training is not given yet but that doesn't mean the rest of the staff is protected, if they don't adhere to the guidelines of policies and the training programmes they can be equally vulnerable as the new employees. Just by conducting this kind of survey in an organization might even give them a feedback about the state of knowledge of the employees regarding the social engineering attack. Another practical method would be to conduct a drill without the employees knowing so that the weak points can be caught easily and a training to cover the mistakes they did can be conducted separately.

V. DESIGN

An effective solution to address the issues of social engineering is needed to address this problem. Figure 11 shows a model as the proactive solution to the issue and this model can be used by organizations as well as individuals to protect themselves against these attacks. This model as depicted can also be applied to the information security policies in organizations to mitigate the risks of the attack. In the proposed model, the organization or the individuals must research on the new and existing social engineering attacks. With the research findings a good information security policy can be developed to give a clear set of guidelines to the personnel on how to disclose sensitive information with clarity on the source which needs the information. Another sub task is to implement good security measures which can be used to protect the information such as installing anti-virus guards to get protected from attachments which was downloaded by phishing emails, security measures can be divided as preventive, detective, corrective, deterrent, recovery and compensating measures. Last but most importantly training must be given to the individuals to identify and respond to such attacks and these trainings are divided into three options, first option is risk which gives an idea about the risk associated with the threat to the personnel. Next, to give them knowledge on how social engineering attacks occur and how they should react to such instances before disclosing the information. Finally, in organizations the training employees must be tested without them knowing, this is called real time testing so that the weak areas of the employees can be identified and the risks can be mitigated. If these measures in preventing social engineering attacks get failed, the failed outcome must be reviewed and the attack must then be researched as shown in the figure 11. This process will be act as a cycle every time a failed defence occurs or a new social engineering threat is surfaced.

VI. CONCLUSION

Social engineering is on the rise, not because it contains less security measures but because the users are not aware about the social engineering tactics that are used by the attackers to gain access to the sensitive information. Social engineering attacks can be divided as Human based attacks and Computer based attacks.("A Proactive Defence to Social Engineering," 2001) Human errors have caused losses in many industries.(Pollock, 2017) Social engineering

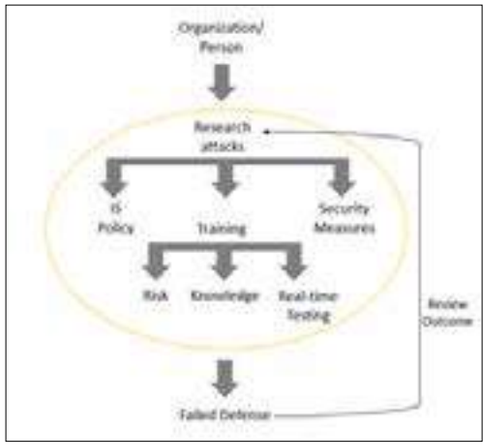


Figure 7. Can the employees be trusted?
Source: Author

attacks such as Phishing attacks which are done by creating different websites are increasing and these sites are from different domains and the average lifetime of those sites are very low.(Cui et al., 2017) If the prevention mechanisms are not enforced to address the social engineering attacks unexpected losses can be surfaced due to the disclosure of sensitive information such as financial loss, loss of public trust and reputation damage. If this situation is critical they might even have to go through legal procedures. To capture the level of awareness of the users from their perspective as well as their organizations, an online survey was distributed and this helped in finding the weak areas of people when dealing with social engineering attacks. The results of the attack give an overall view of the situation and a corresponding solution is proposed to prevent the attack but this does not cover any aspect in detecting the social engineering attacks or responding to the attacks. To mitigate this risk, this research paper has addressed the organizations and individuals with an attack prevention model named as the 'Proactive model A' as a solution which can even be included in the information security policies of the organizations. The aim of this research is to provide a solution to mitigate the risk of the social engineering threats for individuals and organizations. Further advancements include the expansion of the model to Proactive model along with the advancement of threats and risks associated with the issue.

REFERENCES

6 Security Threats to Look Out for in 2018, 2017. . Log. Front LLC.

A Proactive Defence to Social Engineering, 2001. 7.

Banu, D.M.N., Banu, S.M., 2013. A Comprehensive Study of Phishing Attacks 4, 4.

Cazier, J.A., Botelho, C.M., 2007. Social Engineering's Threat to Public Privacy, in: The 6th Security Conference, Las Vegas, Nevada: The Information Institute.

Cert UK, 2015. An introduction to social engineering.

Cui, Q., Jourdan, G.-V., Bochmann, G.V., Couturier, R., Onut, I.-V., 2017. Tracking Phishing Attacks Over Time. ACM Press, pp. 667–676. https://doi.org/10.1145/3038912.3052654

Krombholz, K., Hobel, H., Huber, M., Weippl, E., 2015. Advanced social engineering attacks. J. Inf. Secur. Appl. 22, 113–122.

Luco, D., 2013. The Art of Social Engineering.pdf. ASA Institute for Risk & Innovation.

metasploit, n.d. Best Practices for Social Engineering Attacks.

Mouton, F., Leenen, L., Venter, H.S., 2015. Social Engineering Attack Detection Model: SEADMv2. IEEE, pp. 216–223. https://doi.org/10.1109/CW.2015.52

Pollock, T., 2017. Reducing human error in cyber security using the Human Factors Analysis Classification System (HFACS). 15.

Smith, A., Papadaki, M., Furnell, S.M., 2009. Improving awareness of social engineering attacks, in: IFIP World Conference on Information Security Education. Springer, pp. 249–256.

Social Engineering: A Scheme as Old as Time, 2018. . Secur. Intell.

Social Engineering Attacks: Common Techniques & How to Prevent an Attack [WWW Document], 2015. . Digit. Guard. URL https : // digitalguardian. com / blog / social - engineering - attacks - common - techniques - how -prevent-attack (accessed 5.7.18).

Streeter, D.C., n.d. The Effect of Human Error on Modern Security Breaches 6.

Suganya, V., 2016. A Review on Phishing Attacks and Various Anti Phishing Techniques. Int. J. Comput. Appl. 139, 20–23. https://doi.org/10.5120/ijca2016909084

PERSONALIZED TRAVEL SPOT RECOMMENDATION AND GUIDANCE SYSTEM FOR SRI LANKAN TOURISTS

C Shiranthika[1], N Premakumara, JP Weerawarnakula, H Lakmal, S Fernando, and S Sumathipala

Faculty of Information Technology, University of Moratuwa, Sri Lanka

[1] *chamanijks2@gmail.com*

Abstract- Tourism in Sri Lanka is an evolving field which is significantly influencing the development of the country. With the rapid advancement of Affective computing and its' diverse paths where applications are being implemented by facilitating user needs and emotions, tourism has become one of the prominent fields to provide a comprehensive analysis of useful inclination in specific travel spots based on the user interests and emotions. Traditional tourism methodologies where a travel guide guides on a tourists' journey has nowadays become an old fashion where the tourist himself has innovative applications which provide a guide in almost all the areas in his journey beginning to end. This study proposes a solution where tourist gets a personalized recommendation on travel spots to visit, a summary of the recommended travel spots with a native language translation facility and a translating system to translate landmarks displayed on travel spots such as notice boards and signboards into their native language. Our system divided into four components focusing on (a) profiling users, (b) identifying user locations and travel spots, (c) extracting user reviews about travel spots, summarize and analyse sentiments levels and (d) identifying landmarks displayed in travel spots and translate them into traveller's native language. This approach makes ease traveller's life providing personalized recommendations based on collaborative and content filtering approaches.

Keywords- Personalized recommendations, Travel spot, Sentiment analysis

I. INTRODUCTION

Tourism can be considered as an evolving field in Sri Lanka which significantly influences on the development of the country. It is the practice of touring, attracting, accommodating and entertaining tourists. With the advancement of Information Technology and mobile computing, several innovative ideas have been implemented to facilitate a better experience in the tourist's journey. While the world is rushing under technological enhancements and English becoming the universal language, development of applications in multiple languages will help tourists to achieve more personalized service. In the tourism industry, it is much necessary to provide personalized services to the tourists. When visiting several locations tourists may be much amused if they get the places they most prefer to visit and watch. When tourists have lots of options to choose from which might make them confused in selecting the best possible and or most suitable place to visit, it is essential to filter the information and personalize the choices for the use of each specific user. As a tourist most of the times, it is really confusing to decide where to go and to select among a large number of possible destinations which may also be unknown and unfamiliar. Hence, information retrieval and decision support systems are widely recognized as a valuable context in the tourism domain (B.Rieder, 2013).

Most of the times tourists do not get the full knowledge from the tourist guides and information displayed on the locations. Moreover, in case of unavailability of a tourist guide, they will find difficulties in understanding essential notices displayed on boards such as "හිස් වැසුම් පාවහන් ගලවා ඇතුළු වන්න" "රථගාන ඉදිරියෙන්" etc. Sometimes they will have to follow a multi way process of access internet, search the location, get the details and translate them using an online translation mechanism or ask from another person about displays on boards. Therefore,

development of an application for personalized location recommendation with instant location detail translation is seen as a strong need in the tourism industry.

As the initial step, we consider an approach to recommend similar locations for the places where a tourist has visited using a content-based recommendation approach and we generate an overall opinion with review summary about the location based on the reviews extracting from web documents. System is comprised of four main modules as Profiling module, Location details extraction and summarization module, Review summarization and Sentiment analysis module and Image to text conversion and Native language translation of location details module.

The paper is organized as follows. The first section gives an overview of the existing approaches in travel spot recommendation systems, summarization modules, sentiment analysing modules and translation systems. The second section will describe the overall architecture of the system to be developed. The third section will individually describe the four main modules of the system. Experimental details regarding the implementation will be forwarded to the fourth section. Finally, the paper concluded by indicating future works and new areas to be focused on the tourist guiding applications.

II. METHODOLOGY

This section discusses previously mentioned four main modules of the overall system.

A. Profiling Module

This module consists with the tasks of data extraction from social media, data pre-processing, extracting visual data of a photo, extracting demographic data of the travellers, and unsupervised learning approach of analysing data, the recommendation of the next predictable location and finally evaluation of the results. Here these individual tasks have been explored in the literature separately.

a) Data extraction from social media

Social media is one of the superior ways to get details about places, destinations and also people. This information could be used to get a prior knowledge about a particular place or areas, especially for inexperienced travellers. Collecting details from the geo tagged photos such as the photo URL, location, publisher or the traveller who uploaded it, comments of users, total likes, etc. give many details about destinations and travel experiences. Therefore combining these social traveling experiences associated with users' profile provides the ability to explore the wisdom to have different types of effect and personalized travel experiences.

Literature proves the usage of several methodologies for mining of data from Facebook. NameGenWeb originated at the Oxford Internet Institute, Social Network Importer; a plugin for the NodeXL network analysis and visualization toolkit, Netvizz application are some commonly used Facebook data extractors (B.Rieder, 2013). Despite that Graph API is yet another popular HTTP based API of low level that can be used to query data, post new feeds, handling advertisements, uploading graphic data, and a variety of other needful tasks (A. Gupta, 2017).

b) Data pre-processing

Data analysis in social media is challenging because of the availability of vast volume of data, informal and short messages on social media networks and use of media such as images apart from the usual textual messages to express information by the users. Moreover, availability of noises, emotions will be a considerable challenge in social media. There people's comments written in their native language, punctuations, stop words, URLs, emotions will be removed and convert into a uniform format (J. Han and M. Kamber, 2011).

Therefore, it is needed to handle these incomplete, noisy and inconsistent data to provide quality and structured data where can be used as the inputs for the other modules. There are four major tasks in data pre-processing (J. Han and M. Kamber, 2011) as Data cleaning, Data integration, Data transformation and Data reduction

Under the data cleaning process, it is needed to handle the missing values, noise data and the inconsistent data. To handle missing values, approaches like ignoring the tuple, filling the missing values manually, using a global constant to fill the missing values, using attribute means to fill the missing values are used. To handle the noisy data, data will be smoothed via binning methods, clustering approaches and combining computer and human inspection (J. Han and M. Kamber, 2011).

In the data integration process, the data which have extracted and pre-processed from the different data sources are integrated together to a data store (J. Han and M. Kamber, 2011).

In the data transformation process, the data is consolidated into appropriate forms which can be used for data mining. It involves normalization, smoothing and aggregation.

In the final step of data reduction data cube aggregation, dimension reduction, data compression, Numerosity reduction, discretization and concept hierarchy generation will be done.

c) Extracting visual data of a photo

We focused on the extraction of visual data of the uploaded photos to seek which category the photo is belonging. Literature proves the availability of techniques such as Google cloud vision API, IBM Vision recognition API, Cloud sight API and Microsoft Computer Vision API. Microsoft Computer Vision API, photos assigned 86 different categories like outdoor, outdoor water, outdoor mountain, indoor, sky object, people, etc. with a score depicting on how much probability the photo falls into that category.

d) Unsupervised learning approach of analysing data

Our approach, we assumed that one data point probably belongs to more than one cluster. In our dataset both numerical data as well as categorical data is available which is comprising of post data extracted from Facebook, users' demographic data and the visual data extracted from computer vision API. General clustering approaches include representative based clustering methods such as K-Means algorithm, hierarchical clustering methods such as agglomerative hierarchical clustering, density-based clustering methods such as DBSCAN and Spectral & graph clustering methods (Anon., n.d.). Literature also proves the use of K-Modes algorithm, Squeezer, LIMBO, GAClust, Cobweb algorithm, STIRR, ROCK, CLICK, CACTUS, COOLCAT , CLOPE, etc. in the categorical data clustering[4]. Our approach used the label encoding method to transform the categorical data to numerical data and then apply K-Means clustering approach.

e) Recommendation of the next predictable location

Nowadays in the tourism domain recommendations based on unique personalized factors is a prominent feature which deals with recommending specific items such as restaurants, hotels, sports, activities, accommodations and packaged tour plans also effectively and efficiently. Many researchers have gone through collaborative filtering, content-based filtering and also hybrid filtering approaches when they are to track user's preferences from social networks (A. Gupta, 2017) (Lee, 2017) (Z. Yao, 2016). Based on memory recommender systems use K-Nearest Neighbours algorithm (K-NN) for predicting the preferences of users. Firstly they will identify the nearest users similar to a particular user, obtain the data of the user, and by using a weighting method, recommend items to a user. Based on model recommender systems based on developing a model using machine learning concepts and mainly target the fact how much a user will prefer an item if he did not encounter it before (Lee, 2017) and it outputs more accurate results with more large data than the based on memory methods. Some of the popular model-based approaches are Bayesian Networks, Singular Value Decomposition, and Probabilistic Latent Semantic Analysis (Li, 2017) (Y. Mao, 2018).Researchers have undergone on context-aware recommender systems based on the synergy between soft computing and data mining techniques (G. Fenza, 2011, pp. 131–138).

B. Location details extraction and summarization module

Text summarization methods can be classified into extractive and abstractive (Gupta V, 2010 Aug 20;2(3):258-68.) summarization and extractive (Farshad Kyoomarsi, 2008) summarization methods which focus on selecting essential sentences, paragraphs, etc. from the original document and concatenating them into a shorter form. The importance of sentences is decides based on statistical and linguistic features of sentences.

The abstractive summarization method (Radev, 2004) (Romacker, 2001) consists of understanding original text and re-telling it in the fewer word. It uses scientific methods to examine and interpret the text. To find the new concepts and expressions to best describe it by generating a new shorter text that conveys the most important information from the original text document.

Here we mainly focus on the extractive summarization method. It has two steps as pre-processing step and processing step. Pre-processing is a structured representation of the original text and includes sentences boundary identification, Stop word elimination and Stemming. In Processing step, features influencing the relevance of sentences are decided and calculated and then weights are assigned to these features using weight learning method. The final score of each sentence is determined using feature-weight equation. Top-ranked sentences are selected from final summary. There are three features in the extractive summarization as,

1. Title word feature: Sentences containing a word that appears in the title are also indicative of the theme of the document. These sentences are having higher chances of including in summary.
2. Sentence location feature: Usually first and last sentence of the first and last paragraph of a text document are more critical and are having higher chances to be included in the summary.
3. Sentence Length feature: Huge and concise sentences are usually not included in the summary.

Extractive summarizers (Vishal Gupta, AUGUST 2009) (H. Khosravi, n.d.) aim at picking out the most relevant sentences in the document while also maintaining a low redundancy in summary.

a) LSA Method for extractive summarization

Singular value decomposition (SVD) (H. Khosravi, n.d.) is a potent mathematical tool that can find principal orthogonal dimensions of multidimensional data. It has applications in many areas and is known by different names: Karhunent-Loeve transform in image processes and Latent semantic analysis (LSA) in text processing it gets this name LSA because SVD applied to document word matrices, group documents that are semantically related to each other even when they do not share common words.

Words that usually occur in related contexts are also related in the same singular space. This method can be applied to extract the topic-word and content-sentences from the document. The advantage of using LSA vectors for summarization rather than the word vectors is that conceptual relations as represented in the human brain are automatically captured in the LSA while using word vectors without the LSA transformation requires the design of specific methods to derive conceptual relations. Since SVD finds the principal and mutually orthogonal dimensions of the sentence vectors, picking out a representative sentence from each of the dimensions ensures non- redundancy. It is to be noted that this property applies only to data that has key dimensions inherently. However, LSA would probably work since most of the text data have such principal dimensions owing to the variety of topics it addresses.

C. Review Summarization and Sentiment analysis module

The reviews have become a new way of expressing an individual's opinion through which people openly express their views on various things. Opinions are also valuable when someone wants to hear the other's viewpoint before make a decision. Analysis of these opinions into different classes become a key factor in decision making (Ragha, 2012).So to create a summary of these opinions/reviews, we have to extract topics from whole opinions. Sentiment Analysis is procedure of identifying a selected text or writing is whether positive, negative or neutral. There are two types of textual information as facts and opinions. Reviews are kind of opinions which related to a location, place, person etc. Opinions reflect the people's viewpoint about a thing (Ragha, 2012).

a) Summarization

1) Probabilistic latent semantic indexing technique (PLSI)

PLSI is a maximum likelihood estimation approach.. Documents create a specific distribution of topics p(z|d), Topics produce a specific distribution of word usage p(w|z).Then the probability of generating a document d is,

$$P(w_1 \ldots w_{N_d}) = \prod_{i=1}^{N_d} \sum_{z=1}^{K} P(w_i \mid z) P(z \mid d)$$ (Croft, 2006)

2) cluster-based retrieval

It is also called a mixture of unigram models. In the language modelling framework, the cluster-based topic models were used to smooth the probabilities in the document model (J. Yao, Jan. 2018.). In this cluster-based retrieval model, it is assumed that all documents which use to generate topic models fall into a finite set of K clusters (topics). Documents in each cluster discuss a particular topic z, and each topic z is associated with a multinomial distribution P(w|z) over the vocabulary (Croft, 2006). This model generates following processes on documents;

Select a topic z from a multinomial distribution with parameter

For i = 1..., pick word i w from topic z with probability

P(i w |z)

The probability of generating a document d is using cluster model is;

$$P(w_1 \ldots w_{N_d}) = \sum_{z=1}^{K} P(z) \prod_{i=1}^{N_d} P(w_i \mid z)$$ (Croft, 2006)

One of the parameter estimation methods for the mixture of unigrams model is to cluster documents in the collection into K groups and then use a maximum likelihood estimate a topic model P(w|z) for each cluster. They incorporated the Cluster information into language models as smoothing.

$$P(w \mid D) = \frac{N_d}{N_d + \mu} P_{ML}(w \mid D) + \left(1 - \frac{N_d}{N_d + \mu}\right) P(w \mid cluster)$$

(Croft, 2006)

b) Sentiment Analysis of Reviews

After extracting a review based on a specific place, the application would have an ability to describe it as a positive, negative or neutral. By using this, another tourist can determine whether to that place or not. Researchers describe two main types of approaches for sentiment classification. Semantic orientation method based on a PMI-IR algorithm which combining point mutual information (PMI) and information retrieval (IR). PMI is an association scale between a feature (word) and a category, not between a document and a category. Here X represents the occurrence of a word, and Y represents the occurrence of a category.

Pmi(x, y) = log (P(x, y) / P(x) .P(y)) (J. Wood, 2017)

c) Sentiment analysis

1) Semantic orientation method

Semantic orientation (SO) represents their output using two reference words pair called "excellent" and "poor." The SO of a phrase is retrieved as the mutual information between the given phrase and the word "excellent" minus the mutual information between the given phrase and the word "poor." The equation for that is as below (Croft, 2006).

$$P(w \mid D) = \frac{N_d}{N_d + \mu} P_{ML}(w \mid D) + \left(1 - \frac{N_d}{N_d + \mu}\right) P(w \mid cluster)$$

(Ragha, 2012)

A SO of a review calculated by getting average of SO values of all extracted phrases in the review. This process describes a kind of threshold which use to determine positive and negative of a review (Ragha, 2012). In traditional sentiment analysis methods typically try to extract the overall sentiment revealed in a sentence or document, either positive or negative, or somewhere in between. However, a disadvantage of these methods is texted where a loss of information can also occur. To overcome those issues have to use sentiment analysis method which uses well-trained training sets (Ragha, 2012).

D. Image to text conversion and Native language translation of location details module

This part deals with the images with Sinhala texts captured via mobile devices or uploaded through some media. The primary objective of this module is to capture Sinhala texts and convert it to a character set which can be later used for native language translation purpose.

In Sinhala language, there are 16 vowels, 2 semi-consonants, 40 consonants and 13 consonant modifiers. With the combination of consonant modifiers, a single character can take a large number of shape combination. Also, most of the Sinhala characters take curve shapes, which makes them harder to recognize that most of the other languages. In Sinhala character recognition, we can see most researchers proposed same kind of phases with different or same types of processing techniques. Those phases can be recognize as pre-processing, character segmentation, feature extraction and character recognition.

In (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) and (Dharmapala, 2017), some techniques which have been used have the same final objective but different ways to achieve it. For binarization, in (Dharmapala, 2017) they have proposed the adaptive thresholding technique

while (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) prosing three different types of techniques to achieve the binarized image. In first techniques the sorted gray levels of the image in ascending order and take the maximum gray value of the first quarter as the cut-off value for foreground and background pixels. Second techniques are using a 3x3 kernel which calculate the number of '0' valued pixels and if the cunt is less than threshold value change the value to 255.

For the character, segmentation step both (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) and (Dharmapala, 2017) used vertical and horizontal projection profiles to segment the text lines and non-overlapped and non-touching characters. For the touching characters (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) proposed to categorize segmented characters into different categories such as overlapping, touching, connecting and intersecting based on the average character width. Then to identify touching characters they used a 3x3 kernel which count the discontinuities in lines in characters and by the count they assume the number of characters in segmented part. If there are no discontinuities, they put them into other 3 categories which used "water reservoir" concept to segment the characters. In (Dharmapala, 2017) they proposed to identify the contours of the characters by using OpenCV function named findContours () and redraw the characters separately in canvas.

Character recognition is done by the single Kohenen artificial neural network with 32x32 input neurons and single output neuron in (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) which classify character into one of the fourth predefined character groups. Then depending on the characters groups and using a dictionary most proper word will be generated. In (Dharmapala, 2017), they have used 3 separate neural networks for each zone in character. Then depending on the probability values they produce which indicated relativity with each zone, character can be matched with map of all possible combination of characters.

III. SYSTERM ARCHITECTURE

There are two leading data stores as the Facebook uploaded photo data store and data store with essential notices displayed on travel spots. The Facebook photo data store consists of the user uploaded photos and details

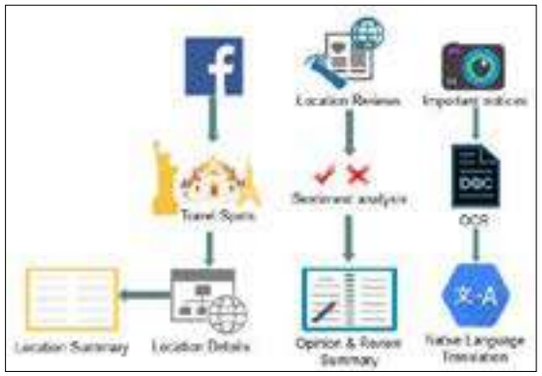


Figure 1. System architecture.

from two Facebook travel groups such as "EnigmaTest" and "Travel Guide Sri Lanka." Post data (publisher name, photo URL, post title. Location tagged) and user demographic data (marital status, gender, hometown, past visited locations) will be sent to the similar location recommendation module. The location outputted will be sent to the location detail extractor module and details of that location will be extracted from the websites. Those will be sent to the summarization module and will generate a location detail summary which will be sent to native language translation module. The comments extracting from the Facebook posts and extracted reviews of locations or travel spots by users from websites will be then sent to the location reviewer where a review summary for the location will be generated through sentiment analysis. That review summary will also be sent to the native language translation module.

IV. APPROACH THROUGH INDIVIDUAL MODULES

A. Profiling module

Main data source in this module is the Facebook uploading photos to selected two tourist groups. There post data was extracted using Graph API for developers assisted by Facebook platform. Photo URL of the post was taken and using Computer Vision API with Python, the category it belonging along with the category score was taken. For a specific user, his demographic data such as relationship status, hometown, age, gender and check-ins were extracted using Graph API.

Group id of the two Facebook groups were gained first.

-Query:GET->/v2.11/EnigmaTest?fields=id, name

-Query:GET->/v2.11/TravelGuideSriLanka?fields=id, name

Post data was extracted via,

-Query:

3054729366178 17?fileds=feed{full_Picture,Place,from,comments{message},message,likes.limit(0).summary(true)}

Here data was normalized, tailored by renaming columns properly, dropping unwanted columns and will clean all the null values. Here we used Scikit learn with Pandas library and done in Jupyter IDE.

Users 'demographic data was extracted via,

-Query:

371223593034449? Fields=id, name, gender, relationship_status, hometown, tagged_places {place {name}}

After we categorized the users based on the past visited locations or check-ins. 18 different locations were identified and we divided the locations in to 5 user types we identified as follows.

We developed a neural network for the user classification purpose and for each user a numerical index was generated based on his check-ins where its numbers are arranged according to the ascending order of his preferences. Here we used Keras library and Tensorflow as the backend.

E.g.:- If the number generated is 325:- 3 denotes Observer, 2 denotes Nature lover, 5 denotes High traveller. Thus from his past check-ins his preferences are in ascending order as above.

Thus the data row generated will be as follows.

Publishers' name, User category (Number we generated), Category score, Category, Location tagged, relationship status, gender, hometown

Here underlined data will be taken for the clustering approach. It was done via K-Means clustering approach.

Table 1. Classification of users with location types.

Location type	User type	Number assigned
Religious locations	Religious person	1
Waterfalls, Watebodies, Beaches, Mountains, Forests	Nature lover	2
Parks, Forts, Monuments, Gardens, Gardens, Markets	Observer	3
Dams and Bridges, Lighthouses , Powerplants	Building explorer	4
Islands, Countries, Scenery Towns, Holiday Resorts	Long distance traveler	5

There we used scikit learns label encoding approach to convert the category and location tagged which are categorical in nature, in to numerical values. In the clustering process we used the Elbow method to find the most optimal number of clusters which experimentally 15 was gained as the optimum number.

Users was clustered likewise and to suggest next locations for a particular user, we used the locations visited by the similar users to him who are in the same cluster.

B. Location details extraction and summarization module

Here we applied the concept of making the location summary using NLTK library and fuzzy logy and applied on to multiple documents to extract specific sentences. The algorithm we used here extract one or more sentences that cover the main topics of the original document using the idea that, if a sentence contains the most recurrent words in the text, it most probably covers most of the topics of the text.

Figure 2. Image to Text Conversion main module

Here we used the concept of frequency summarizer. It tokenizes the input into sentences then computes the frequency map of the words. Then the frequency map is filtered in order to ignore very low frequency (frequency<0.3) and highly frequent word (frequency>0.9) in order to discard the noisy words such as determiners and to discard words such as stop words (a, the, an, that, one etc.) which do not contain much information. And finally the sentences are ranked according to the frequency of the words they contain and the top sentences are selected for the final summary. The simplest method to use frequency of words as indicators of importance is word probability. The probability of a word w is determined as the number of occurrences of the word, divided by the number of all words in the input. Since it is not accurate if we get summary using only NLTK we use fuzzy logy technology to improve the accuracy. Here we first set the number of sentences in the summary as 5, but we facilitates the user to get as his preferred number of sentences.

C. Review Summarization and Sentiment analysis module

Input data to this module are reviews (comments) about locations. Output is a summary of reviews and an overall sentiment analysis of them. This process contain several sub process as data gathering, topic modelling, summarization and sentiment analysis. Data was gathered by web scrapping techniques because it is easy to get only the required data from websites by specifying the html tag which data is located. Data was pre-processed by using the enchant module which have an English words corpus by removing stop words and other languages texts. To create a summary from these opinions/reviews we used topic modelling approach and extract text which have highest frequency among all the document data. Here we used LDA based topic modelling approach. Here a particular topic is defined by a cluster of words with each word in the cluster having a probability of occurrence for the given topic, and different topics have their respective clusters of words along with corresponding probabilities.

After summarizing sentiment analysis module was implemented to obtain sentiments about that place according to reviews. By adding this module it will help to tourists to make decision about a place whether to visit or not. Here we used NLP based sentiment classification algorithm which would be able to design a well-trained training set by adding weights to the words. The trained training sets can easily filter the attributes as per the user requirements. And also we used the Bayesian algorithm to classify the sentiments as positive or negative which is using a word dictionary and training data sets to analysing process. Through that data lose will be reduced.

D. Image to text conversion and Native language translation of location details module

In this module we hope to follow the same four phases mentioned in referenced researches (M.L.M Karunanayaka, 29 Nov-01 Dec 2004) (Dharmapala, 2017) with one additional phase. So in our approach we separate this module into pre-processing, character segmentation, feature extraction, character recognition and word correction.

In pre-processing phase we decided to use Contrast Limited Adaptive Histogram Equalization (CLAHE) as the contrast equalization method. To remove noise in the image we use Non-local means denoising method and mean filter. For binarization step Otsus's algorithm is used which is kind of an adaptive thresholding which consider image pixel as background and foreground depending on the threshold. For skew correction process Probabilistic Hough Transformation is decided to use which use Hough transform lines to correct the angles.

In Character segmentation phase we decided to use the Horizontal Projection Profile to segment text lines and vertical projection profile to segment words and characters. For the overlapping and touching characters, contours will be considered and image will be redrawn in separate canvases. Contour recognition will be performed for both overlapping and touching characters and characters with modifiers. So when we segment a character with modifier we have to perform contour recognition and redraw steps again to segment character and modifier.

In feature extraction we used the horizontal and vertical projection profiles of the three zones in image named as upper, middle and lower which take place in first one third from top of the image, second one third form middle of the image and last one third from bottom of the character image. Vertical and horizontal projection profiles are for two parts from each zones which divided horizontally for horizontal projection profile and vertically for vertical projection profile. Then the middle zones pixel values is taken as separate vector.

In Character recognition phase above feature vectors are processed through three separate neural networks which process upper, lower and middle feature vectors separately. Out will be a probability value that express how much given feature vector is related to given label. Then by considering joint probability for all three zones and matching joint probability with map of all possible character combination, a character will be generated. By assembling those character, we can generate a word.

In word correction phase we correct the generated word by using a verified Sinhala word corpus and using Bayes theorem as basic approach.

V. RESULTS

We used Python as the programming language and used ionic as the front end development tool. Here we used Mongo DB in the storing of data extracted by Facebook and processed data.

From the Location recommendation module a location based on a users' past visited locations and on his uploaded photos will be generated. From the Location details extraction and summarization module a summary will be generated for a location entered by the user. From the Review summarization and Sentiment analysis module a review summary for a particular location will be generated with an overall sentiment analysis. From the Image to text conversion and Native language translation of location details module user can get the translated details of the Sinhala notices displayed in landmarks.

VI. EVALUATION OF RESULTS

As per the evaluation, we prepared a user evaluation form to be filled by 50 selected users. There 7 criteria related to the system performance were identified and users were told to put a tick in front of the criteria whether it is excellent, good, satisfactory, fair or worse. Criteria and the results obtained are as follows.

Criteria A - Recommend locations that you like most

Criteria B - The user types suggested matches you

Criteria C - Location Detail summary is accurate

Criteria D - Location detail summary contain enough details

Criteria E - Review summary generated about a location is accurate and meaningful

Criteria F - Sentiment analysis about locations is accurate

Criteria G - Understanding of translated recognized word in captured notices

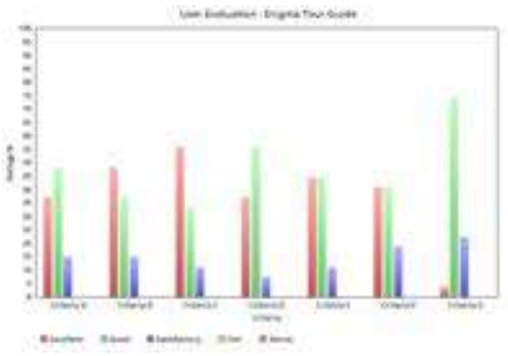


Figure 3. User Evaluation results

VI. CONCLUSION

In this paper it is being discussed the use of our system as a smart guiding system, which will help travellers to customize their journey accordingly. We have tried to provide a system which will help tourists to freely travel and find travel spots where they do not see a need of a third party to guide them. We have undergone several used approaches, techniques and methodologies which researchers have used in the similar tasks to our tasks. Through our system, it was focused to get the most applicable technique for different tasks while improving the accuracy of the final system.

ACKNOWLEDGMENT

Authors would specially thank Dr. Sagara Sumathipala and Dr. Subha Fernando who gave a massive guidance and support in understanding the research domain. Acknowledgements also go to all the lecturers of Faculty of Information Technology, University of Moratuwa for their assistance throughout the research and the system implementation.

REFERENCES

A. Gupta, H. B. P. S. a. P. K. S., 2017. *Facebook Based Choice Filtering*,. s.l., Advance Computing Conference (IACC), 2017 IEEE 7th International, 2017, pp. 875–879..

Anon., n.d. Clustering categorical data streams.

Anon., n.d. *En.wikipedia.org(2017). Pointwise mutual information*.[Online]Available at: https://en.wikipedia.org/wiki/Pointwise_mutual_information[Accessed 21 Dec 2017].

B.Rieder, 2013. *Studying Facebook via data extraction: the Netvizz application*. s.l., 5th annual ACM web science conference.

Croft, X. W. a. W. B., 2006. *LDA-based document models for ad-hoc retrieval*. s.l., 29th annual international ACM SIGIR conference on Research and development in information retrieval, 2006, pp. 178–185..

Dharmapala, K. A. K. N. D. W. W. P. M. V. C. C. P. R. U. K. A. U. &. R., 2017. Sinhala Handwriting Recognition Mechanism Using Zone Based Feature Extraction..

Farshad Kyoomarsi, H. K. E. E. a. P. K. D., 2008. *Optimizing Text Summsarization Based on Fuzzy Logic*. UK, Seventh IEEE/ACIS International Conference on Computer and Information Science, IEEE, University of Shahid Bahonar Kerman, UK, 347-352.

G. Fenza, E. F. D. F. a. V. L., 2011, pp. 131–138. *[9] G. Fenza, EA hybrid context aware system for tourist guidance based on collaborative filtering*. s.l., Fuzzy Systems (FUZZ), 2011 IEEE International Conference on, 2011.

Gupta V, L. G., 2010 Aug 20;2(3):258-68.. *A survey of text summarization extractive techniques*, s.l.: Journal of emerging technologies in web intelligence.

H. Khosravi, E. E. F. K. a. P. K. D., n.d. Optimizing Text Summarization Based on Fuzzy Logic. *Computer and Information Science Studies in Computational Intelligence,* p. 121–130..

J. Han and M. Kamber, D. m. c. a. t. 3. e. B. M. E. 2., 2011. *Data mining: concepts and techniques*. 3rd ed. ed. Burlington: MA: Elsevier, 2011..

J. Wood, P. T. W. W. a. C. A., 2017. Source-LDA: Enhancing Probabilistic Topic Models Using Prior Knowledge Sources. p. pp. 411–422..

J. Yao, Y. W. Y. Z. J. S. a. J. Z., Jan. 2018.. Joint Latent Dirichlet Allocation for Social Tags. *IEEE Trans. Multimed*, Volume vol. 20, no. 1, p. pp. 224–237.

Lee, C.-Y. S. a. A. J. T., 2017. Tour recommendations by mining photo sharing social media," Decis. Support Syst., ,. Sep , pp. vol. 101, pp. 28–39.

Li, S. L. a. G., 2017. Personalized Hotel Recommendation based on Social Networks.

M.L.M Karunanayaka, N. K. a. G. W., 29 Nov-01 Dec 2004. *Off Line Sinhala Handwriting Recognition with an Application for Postal City Name Recognition*. s.l., 6th International Information Technology Conference on From Research to Reality, Infotel Lanka Society Colombo, Sri Lanka.

Pal, U. &. D. S., 2003, August. Segmentation of Bangla unconstrained handwritten text. p. 1128.

Radev, G. E. a. D. R., 2004. LexRank: Graph-based Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research, Re-search*, , Volume Vol. 22, pp. pp. 457-479 .

Ragha, T. G. a. L., 2012. *Featured based sentiment classification for hotel reviews using NLP and Bayesian classification*. s.l., Information & Computing Technology (ICCICT), 2012 International Conference on, 2012,p.1-5.

Romacker, U. H. a. M., 2001. The *SYNDIKATE text Knowledge base generator*. ACM, Morristown, NJ, USA, first International conference on Human language technology research, Association for Computational Linguistics .

Turney, P. D., 2002. *Thumbs up or thumbs down?*. s.l., 40th annual meeting on association for computational linguistics, 2002, pp. 417–424..

Vishal Gupta, G. L., AUGUST 2009. A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence,* VOL. 1(NO. 1), pp. 60-76.

Y. Mao, F. Z. L. X. D. Z. a. H. Y. J.-S. P. T.-Y. W. Y. Z. a. L. C. J., 2018. A Bidirectional Collaborative Filtering Recommender System Based on EM Algorithm. *Advances in Smart Vehicular Technology, Transportation, Communication and Applications*, vol. 86(Eds. Cham: Springer International Publishing), pp. pp 265-273.

Z. Yao, Y. F. B. L. Y. L. a. H. X., 2016. [7] *[Z. Yao, Y. Fu, B. POI Recommendation: A Temporal Matching between POI Popularity and User Regularity*. s.l., Data Mining (ICDM),2016 IEEE 16th International Conference on, pp. 549–558..

Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization. In: Chapter VIII.5*, Graphics Gems IV*. *Cambridge:* MA, Academic Press, pp. 474-485.

WEB, MOBILE AND COMPUTER ACCESSIBILITY: ISSUES FACED BY THE SRI LANKAN VISUALLY IMPAIRED COMMUNITY

N Wedasinghe[1], N Sirisoma[2], and APR Wicramarachchi[3]

[1 2]General Sir John Kotelawela University, Ratmalana, Sri Lanka

[3]University of Kelaniya, Sri Lanka

[1]niroshaw@hotmail.com

Abstract- Though Information Technology and Internet provide benefits to their customers, there is still a gap existing between none-differently abled and differently abled users. This gap is known as Disability Digital Divide. When compared to none-differently abled, differently abled users face a disadvantage when accessing these technologies. In Sri Lankan context, there is no proper planning or guidelines to overcome these issues specifically faced by the visually impaired community. Therefore, this research focuses on addressing those issues and finding constructive solutions. The study focuses on three main research questions. Firstly, it identifies the problems and issues faced by visually impaired people when accessing personnel computers, mobiles, Internet and web related technologies. Secondly it concentrates on technological accessibility related issues in relation to these technologies and finally, how the above problems and issues can be rectified. This study was conducted by engaging in in-depth interviews with visually impaired individuals and observing Computer, Web and mobile accessibility issues. Snowball sampling was used and this research directly benefits visually impaired community allowing them to overcome the obstacles, problems and issues they are facing in their day to day life in the context of Information accessibility. Findings indicated that, current websites failing to adhere to Web Accessibility guidelines, difficulties in software accessibility, human perception on technology, financial difficulties to purchase and use of equipment are the major issues. Solutions recommended to overcome such issues and improve the accessibility among the Sri Lankan community include standardization of web and internet facilities, concentration on user friendliness in software development processes, infrastructure development, and financial support for visually impaired people and special training and education on technology with proper guidance.

Keywords- Web Accessibility, Mobile Accessibility, Computer Accessibility, Information Accessibility, Visual Impairment

I. INTRODUCTION

For most people, technology makes things easier. For people with disabilities, technology makes things possible". The often-cited quote by Mary Pat Radabaugh, former Director of IBM National Support Center for Persons with Disabilities, sums up the importance of technology in the process of empowering persons with disabilities (National Council on Disability 1993). The exclusion and marginalization of persons with disabilities is a human rights issue as well as an economic issue for countries. When a significant section of society, estimated at 15% of the world population, faces obstacles in receiving education, transitioning into the labor market, and becoming economically self-sufficient, it does not only undermine their rights and dignity but also adds significantly to a country's welfare burden (WHO and World Bank, 2012).

Though Information Technology and Internet provide more benefits to their customers, but there is still a gap between none-differently abled users and differently abled users. This gap is known as Disability Digital Divide.

Among that disadvantage group according to principle of School for the blind Rathmalana stated that blind and visually impaired individuals are the most marginalized group. This has been caused by a number of conditions such as poverty, self-employment, lack of education, getting discouraged to obtain computer literacy due to lack of support. Hollier (2007) also indicated that most disadvantaged group among the differently abled community is the visually impaired community. Wedasinghe, Wicramarachchi and Sirisoma (2017) disability digital studies related to information accessibility also stated that visual impaired community become the most disadvantaged group among the differently abled community.

However, when considering Sri Lankan context there is no proper planning or guidelines to overcome the disability digital divide issues faced by the visually impaired community in Sri Lanka. Therefore it is necessary to identify current problems they face and find constructive solutions to overcome such current issues in Sri Lanka.

This study focuses on three main research questions such as; What are the problems and issues faced by visually impaired people when accessing personnel computer, mobile, Internet and web related technologies in Sri Lanka?, Are there any technological accessibility related issues in the computers, mobile technology and web related technologies experienced by visually impaired community of Sri Lanka? , How the above problems and issues related in the context of Sri Lanka, can be overcome specifically for vision impaired people?.

The significance of this research can be identified as a new dimension of bridging the disability digital divide in relation to visually impaired community. And it is also contributes to Sri Lankan society by identifying means to overcome the disability digital divide.

II METHODOLOGY

A. Study Design

This research is conducted as a second phase of study. First phase of this study conducted as a survey in relation to differently abled community and the second phase of this study is a qualitative based study. First part of this study has been exploring that out of all differently abled community in Sri Lanka, the most disadvantage community is visual impairment community. In addition to that survey conducted by censer and statistics department of Sri Lanka (2014), indicated that out of all disabled categories 54% of them are suffering from visual impairment. Therefore, this phase focuses to study in-depth interviews and observations to identify the information accessibility related issues faced by the blind community in relation to Sri Lankan context.

B. Study setting

In the first phase of this study has been conducted to identify the Information accessibility related problems and issues faced by the differently abled community in Sri Lanka. Therefore In this phase consider only visually impaired community. This study is setting to conduct in-depth interviews with visually impaired individuals and observations of Computer, Web and mobile accessibility issues among visually impaired community in Sri Lanka.

C. Target Population

The Target population for this study is visually impaired community in Sri Lanka. According to the Department of census and statistics 2014 indicating total visual impaired population in Sri Lanka is 996939 and they are including with blind and partially impairment on vision. However the technology usage by this community is not identified based on the web, mobile and computing.

D. Sample size

Since this is a qualitative research the issue of sampling has little significance as the main aim of most qualitative inquiries is either to explore or describe the diversity in a situation, phenomenon or issue. Qualitative research does not make an attempt to quantify the findings size of the sample is not an important issues. In this research saturation point will be consider. Therefore, in this research we are using non-probability sample until reaching the saturation point.

E. Sampling method and participants

Since this phase is using qualitative approach, sampling method will be used for this study is snowball sampling. Snowball sampling is the process of selecting a sample using networks. To start with, A few individuals in blind or visually impaired organization be selected. They are then asked to identify other people in the group or organization, and the people selected by them become a part of the sample. This process is continuing until the required number or a saturation point has been reached, in terms of the information being sought.

F. Inclusion and Exclusion criteria

Participants were all visual impaired individuals with having different capacity of vision impairments. It is including with the individuals who are totally blind or some form of a vision difficulties. This could be happen from the birth itself or at the part of their life. The participant need to be interesting and awareness of computer, mobile or Internet.

The individuals who do not know about Information accessibility on either one of technology discussed above will be excluded from this research. In addition to this the age group below 10 years and the people who are having eye vision with mental disorders also excluded with this research.

III EXPERIMENTAL DESIGN

In this study, first researcher will be contact initial network through School for the blind Rathmalana. This institute is located near to General Sir John Kotelawala Defence University. Therefore In order to make an initial network will be convenient. The requesting letter to giving permission researcher to collect data form the visual impaired community from the faculty of graduate studies collected and send to principle of the school for the Blind. With the approval from the principal researcher can collect data from the visual impairment community. In addition to that individual who are having visual impairment was interviewed to obtain findings.

All the potential participants will be given an invitation letter with a detail information leaflet written in appropriate lay language [English,Sinhala]. This information leaflet describes the purpose of the study, why they have been chosen, anonymity, risk and benefits of participation. The participants given time to read and understand the information leaflet. Since this target group is blind or vision difficulties, Softcopy of the information leaflet will be given to listen via screen reading software. The participant can be request any form of leaflet. According to the participant choice brail version of the information leaflet can be produce. If the participant is with a dependent this information leaflet can be share with them and discuss before deciding to take part in the study.

After they agree to participate for this study, participant will be contact via phone they would like to participate in the study and the opportunity will be given to answer questions they may have about the study. After agree to take part in the study, a convenient date and time location will be arranged for in-depth interviews. However, if following this discussion the participant would like more time to consider their participation. Follow up call given before start the in-depth interview.

If the participant agrees to take part in-depth interviews, immediately prior to the interview taking place, signed consent obtains from each participant. It made clear to the participant their participation is on a voluntary basis and that they do not have to participate if they do not wish to do so, and can refuse to answer any questions at any time or terminate any point during the interview. At any given point participant can opportunity to as anything about the research and the purpose of asking such questions, detail description about the research.

Participant will be informed of their right to withdraw from the study before or during the survey and up to one month after their date of data collection. If they choose to withdraw and remove the data from the research, those data will be not taken in this research. However it will be made clear to participants that if they choose to withdraw after a period of one month.

A. Data Collection and Planned Investigations

In this research data collected with using in-depth interviews. This in-depth interviews researcher will make sure that the risk of providing this data in minimal. Minimum risk means that the extent of harm or discomfort in the research is not greater than that

ordinarily encountered in their daily life. Avoiding bias of the answers to the relevant questions researcher will be not influence philosophical idea on this blind community. The collected raw data will be editing before analysis and made conclusion. All the data collecting through an in-depth interview recoded in audio format. This will be further listen for clarifications.

B. Data Entry and analysis

The collected recording data will be entered in to the form of text format based on the themes which are aligning to research objectives. Context analysis will be used to analysis the collect data which means to identify the main themes those emerge from the responses given by the interviewees. Following steps followed to coding to descriptive data.

Step 1 : Identify the main themes. Researcher will be carefully go through descriptive given responses by the interviewee to each question to clarify the correct meaning. Different meanings used for the same word identified in this step. This theme used as a basis to analysis the data.

Step 2 : Assign codes to the main themes. This step is follow to identify the number of times a theme has occurred in an interview. This themes will be written and assign a code to it.

Step 3 : Classify responses under the main themes: Having identified the themes, then transcripts of all interviews and classify the responses under the different theme. NVivo qualitative data analysis software used to analysis of data.

Step 4 : Integrate themes and responses into report. Finally all the findings will be integrate and include in the report.

C. Plan for dissemination of Findings

The research findings will be shared through the publicly in the form of a thesis. In addition to that findings disseminate in the form of oral presentation or poster in scientific sessions, academic forums, conferences and papers written to high impact, open access academic journals. Significant findings shared with the Information and communication Technology Authority (ICTA) and contribute towards policy formulation. In addition to that findings shared publicly to improve the awareness and increase support visual impaired community in Sri Lanka.

IV. RESULTS AND DISCUSSION

Problems and issues faced by visually impaired people further discuss (Table 1) based on the three accessibility areas on computer, Mobile and Internet or web.

Table 1. General Problems faced by Visual impaired community in Sri Lanka

Accessibility Type	Problem
Computer	• Screen Reading Software compatibility issues • Not familiar with the Software • Key board arrow keys are not supporting • attitudes toward to computer accessibility • Language fluency
Mobile	• Uneasiness of touchable smartphone • Screen size • Complexity of items
Web	• Expensive to use Internet Data

A. Accessibility of personnel computer

According to the findings identified barriers grouped into three functional categories: barriers to providing computer input, interpreting output, and reading supporting documentation.

The main Computer input barrier for this community is supporting screen reading software is not familiar to some computer users. Some expert users who are daily using computer in their day-to-day life are more convenience of using computers. Problem face by this community is depending on the level of vision impairment. People who are blind are using key board arrow keys. But if the relevant software is not supporting this functionality the people will be difficult and frustrated of using computers. They feel that this machine is not for help them to simplify their life but complicated their work .In the observation it has shown level of usability of computer is related with their attitudes towards Computer accessibility. In addition to that practicing key board to this community is not simple as non- vision impaired or blind individual .Because they have to keep everything in their mind. Language fluency is another issue for accessibility of computer. Many supporting software are in the form of English language and not supporting Sinhala Unicode system.

B. Accessibility of mobile

Findings indicated that some blind people like to use keypad supportive mobile phones. They stated that touch system are not easy for them and difficulty of accessing. But some of them are prefer to use smart phone and the awareness of this technology among them is higher. They are getting much social interaction via this smart phone. But the people using touchable mobile phone are using their personal computers to communicate with new technologies. They are using social media like Facebook, Whatsapp etc. Visual impaired People with partially sighted vision capability are complained that functions available in the mobiles are sometimes complicated to understand.

C. Accessibility of Internet and web related technologies

In the Sri Lankan context Free Wi-Fi is a very rare experience for a visual impaired community. They are stated that expensiveness of getting Internet data from Internet providers are not affordable. This may cause lack of interesting of accessibility of Internet.

D. Technological difficulties of accessibility on computers

Getting licence software for them becomes a dream unless they are funding from a funding donor organization. Their self-finance is not capable to pay licence software. Therefore most of them are using crack version of software. This is an illegal and as a solution they can use open source software. Still they face this software are not familiars and user-friendly for them to use and make them uncomfortable.

Table 2: Technical problems faced by Visual impaired community in Sri Lanka

Accessibility Type	Problem
Computer	• Cost is unaffordable with licence software's and crack versions are using with limited access • Open source screen reading software are not user-friendly and voice is not human voice and users feel un-easiness
Mobile	• Low cost phones are available limited no of function • New Apps are unable to install • Some accessibility features are only available with Apple iPhones but not with other phone types
Web	• Web sites are not supporting screen reading software's • W3C guidelines are not followed by the web developers • Specifically, in Social media Pictures are not in transfer without description

E. Technological difficulties of accessibility on mobile technology

The main problem with them is that they are using old type phones are not providing additional support. But cost is less and affordable them financially. On the other hand many of new Apps are unable to install and some apps are built for only some platforms and not universally support.

F. Technological difficulties of accessibility on web related technologies

When they are using Internet the most critical problem they face is most of the website are not readable. The cause for this issue is when developing websites, they were not used W3C web accessibility Technology Agency of Sri Lanka (ICTA) also already given such guidelines. But

most of the websites are not using these web standards. Therefore, this community is getting disadvantage of using this web access. Another main problem they mention is when they are accessing graphics screen reading software is not transferring emotions and they are not able to understand the picture. Further they mention sometimes just saying three people are stand in a picture. But screen reading software is not mentioned the expressions of this three people. Therefore, visual impaired person is difficult of imagine and understand the way that a person is not having visual impairment identify about the same picture.

V. CONCLUSION AND RECOMMENDATIONS

According to the above identified problems and issues related in the context of Sri Lanka can be overcome specifically for vision impaired people in multiple ways.

Develop screen reading software with user-friendly and give them a proper training for them and make sure they came to satisfactory level and evaluate them will recommend.

Supporting training with computer key board functionalities for computer software will make improvement of the community.

Improvement of English language usability and IT technology usability make together will recommended implementing to improve the accessibility among this disadvantage community.

Sinhala Unicode support validate checking make compulsory for web site at least for the most popular and relevant web sites for visual impaired community.

When it comes to mobile phones double click method could be activated and training them will improve the accessibility of smart phone among this community.

Giving some special rate data packages for this community also recommended to improve the accessibility of this community.

Promote and funding to purchase licence software without illegal copywriting issue make them software accessibility improvement in an ethical environment.

REFERENCES

Hollier S.E.(2007) The Disability Divide: A Study into the Impact of Computing and Internet-related Technologies on People who are Blind or Vision Impaired. Published PhD thesis, Curtin University of Technology

Leporini, B. & Paterno, F. (2008). Applying Web Usability Criteria for Vision-Impaired Users: Does It Really Improve Task Performance? International Journal of Human-Computer Interaction. 24, 17-47.

National Council on Disability(1993) Study on the financing of assistive technology devices and services for individuals with disabilities. A report to the President and the Congress of the United States. http://www.ncd.gov/publications/1993/Mar41993.

Power, C., Freire, A., Petrie, H., Swallow, D. (2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12). 433-442

Raufia B., Feratia M., Zenunia M., Ajdaria J., Ismailia F. (2015). Methods and Techniques of Adaptive Web Accessibility for the Blind and Visually Impaired, Procedia - Social and Behavioral Sciences 195 (2015) 1999 – 2007 , Elsevier

Semaan, B., Tekli, J., Issa, Y.B. Tekli, G., & Chbeir, R. (2013). Toward Enhancing Web Accessibility for Blind Users through the Semantic Web. Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on. 247-256.

Söderström, S., & Ytterhus, B. (2010). The use and non- use of assistive technologies from the world of information and communication technology by visually impaired young people: a walk on the tightrope of peer inclusion. *Disability & Society*, 25 (3),303-315.

Wedasinghe N, Sirisoma N, Wicramarachchi APR (2017),' Building digitally inclusive differently abled community in Sri Lanka: an Information and Communication Technology (ICT) policy framework', 10 th International Research Conference 2017, General Sir John Kotelawala

Wedasinghe, N., & Wicramaarchchi, R. (2014). Web, mobile and computer related

model to bridge the disability digital divide in Sri Lanka. *International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) 2014* (pp. 1-6). Dhaka: IEEE.

World Health Organization (2012) Sri Lanka launches the World Report on Disability. [online] [[Accessed 30 th Nov 2017]available at http://www.who.int/disabilities/media/news/2012/30_05/en/index.html

GPS BASED SAFE LOCATION GUIDING ANDROID SYSTEMS IN CASE OF TSUNAMI

RVC Tharanga1 and RMM Pradeep
Faculty of Computing, Department of Information Technology,
General sir John Kotelawala Defence University
[1]randunuveera.1995@gmail.com

Abstract- Tsunami can be defined as giant waves caused by earthquakes or volcanic eruptions under the sea. Out in the depths of the ocean, tsunami waves do not dramatically increase in height. But as the waves travel inland, waves build up to higher and higher heights as the depth of the ocean decreases. It has been scientifically proved that the speed of tsunami waves depends on ocean depth rather than the distance from the gravity of the wave. This research is to establish a tsunami guidance system and to make the general public aware about such a disaster and to guide them to the nearest security sites. The major problem that has been identified in this research is the impossibility of arriving in a safe place and the lack of knowledge in case of tsunami. The probability of predicting a disaster is very unlikely, thus the resulting damage is immense. Once this scenario is taken in to consideration, this research has identified some problems. Among them, the lack of preparedness for potential tsunamis, the lack of resources and the exotic staff have no idea of reacting and getting warning signals. But one of the major issues of this research is the inability to track Tsunami alerts and arrive at a safe place in the easiest way.

Keywords- Tsunami, easiest way, distance, warning signals

I. INTRODUCTION

The purpose of this project is to alert people living in an uncertain area of the tsunami in southern Sri Lanka and guide them to safe places. According to the pre-survey of people living in the South, research has shown that people living in close proximity to the coastal zone are aware of the safe places that are pre-appointed by the government but those people lack knowledge about the actions to be undertaken. In case of a Tsunami warning, hence people are not aware of the immediate actions that need to be taken.

This project is an introduction to an Android application that must be installed on the smartphone and detects alert alarms sent by tsunami warning systems which have been installed along the coastline in Sri Lanka. What is more, this app shows the nearest security locations that have been named by the government in case of an emergency and the fastest way to access the designated security locations using Global Positioning System (GPS). The purpose of this project is to set up a tsunami warning system and to raise public awareness on this disaster and to guide them to the closest security sites by the fastest route. Ultimately, this can reduce the number of deaths caused by the tsunami and ensure that people are not affected and rest in safe places. (Olson et al., 2011).

II. LITERATURE REVIEW

It is noteworthy that actions and responses taken in the initial minutes of an emergency are critical. These life-threatening events may happen at any moment. Being always prepared and taking precautions can save lives. A call for help to public emergency services that provides full and accurate information by the general public who're affected will help the dispatcher to send the right responders and equipment.

It is worth recalling the research done by Jethro B. de Guzman, Ritz Carlo C. de Guzman, and Engr. Remedios G. Ado on Mobile Emergency Response Application Using

Geolocation for Command Centres (de Guzman et al., 2014). According to their research, this is a combination of a mobile application and a web application to respond to urgent ambulance, fire truck and police requests by people in a certain area or city. The mobile application would detect the user's current location via geolocation and send it to the deployed web application in a command centre with the name, age, mobile number and location of the user for easy dispatch of emergency units.

Mao Chongyuan and Fu Qiang have done a research about APPLICATION RESEARCH OF TSUNAMI MONITORING AND FORECASTING USING DATA MINING TECHNIQUES (Chongyuan and Qiang, n.d.). The following result is borne out by this research on the tsunami triggered by earthquakes, it is difficult to detect and of great danger. As the tsunami triggered by an earthquake on the high seas is spreading rapidly, it is difficult to effectively monitor and forecast in an effective and timely manner. However, the trajectory and impact of the tsunami follow certain rules that are not obvious.

Marco Romano, Teresa Onorati, Ignacio Aedo and Paloma Diaz have done an article regarding designing Mobile Applications for Emergency Response: Citizens Acting as Human Sensors (Romano et al., 2016). According to their article when an emergency occurs, citizens can be a useful support for the operation centres involved in intervention activities. As witnesses to a crisis, people can initially share up-to-date and detailed information about what's going on.

J. Wachter, A. Babeyko, J. Fleischer, R. Haner, M. Hammitzsch, A. Kloth have done a research about Development of tsunami early warning systems and future challenges (Wächter et al., 2012). According to their research Encouraged by and integrated in the general development of information and communication technologies (ICT), the evolution of the tsunami warning systems (TWS) shows a significant development of the architectures of seismic systems fluted to multisensory using additional sensors for tsunami detection in the ocean. Currently, the initial implementation of the regional tsunami warning infrastructure indicates a new phase in the development of TWS. The research underlined that a new generation of TWS should not only be able to perform multi-sensor monitoring for tsunami detection.

M. Di RisioG. has done a research regarding Algorithms for automatic, real-time tsunami detection in wind-wave measurements: using strategies and practical aspects (Di Risio and Beltrami, 2014). Based on the findings of research, the authors identification of the ability not only to detect, automatically and in real time, a tsunami propagating on a WWG site, but also to characterize its waveform are notable. With regard to this method, the tests show that it is capable of almost perfect characterization of the tsunami waveform, at least, in the case of, tsunamis generated by an earthquake.

Tan Juan has done a research regarding Risk Assessment of Computer Network Security in Banks (Juan, 2016). According to the research, in response to the media report of holes in the banking system and the credit crisis banks, the study highlighted the banks should improve the management of the security of its computer system and increase security levels. Priority should be given to assessing the risks associated with the security of the computer system, so as to ensure the normal operation of the bank's business.

III. METHODOLOGY

Android, GPS accompanied by GIS technology can be used to develop an Android system that works as a guide and directs people from the current location to a place of safety, useful for people traveling near the coastal area can be warned if such warning is published

The main result of this research is to develop an android solution to show the shortest way to reach the tsunami safety site. Depending on the user of the currently hosted place, the system will indicate the possible security locations identified by the government. Depending on the security locations, the user can choose the nearest security location. If people select the security location as much as possible, the system will guide people to the location.

To study the key issue, there is little action to be taken. The main objective of the survey was to understand the knowledge of the place of safety in case of tsunami of the community. Firstly, the research identified the problem. To identify the problem, the above research has already been done. After that, related research distributed an online survey to obtain data from the community. A questionnaire was prepared and distributed among people who live in the Southern Province and other provinces. According to the survey, several search results were found. After mentioning the results of the survey:

A. Lack of preparation for a potential Tsunami.

The author conducted an online survey so that the majority of survey participants represented the Southern province and the majority of them had experienced the tsunami disaster in 2004. Although he is already 12 years old, the author noticed some persistent issues that have improved the requirement for a sophisticated Android system to communicate tsunami warning alarms. According to the author, it was noted that the majority of them had not participate in the government-led tsunami training programs and that the majority of them do not seem too enthusiastic about this training as such training would divert their daily activities.

B. Inadequacy of Resources

Disaster warning alarm systems require high technological equipment, infrastructure, knowledgeable staff or expensive services. This research came across certain instances where the basics of such a sophisticated alarming system would cause high cost but ultimately the general public don't receive the message of a disaster communicated on time.

C. No idea of reacting and getting warning signals.

It was identified that the tourists or whoever visit down south as strangers had no idea of the warning alarm system as the pre-tsunami training was only given aiming a particular target group. It is observed that the need for an instant warning system for anyone who are close to the disaster area is seen as an urgent requirement.

According to above findings, this project tries to implement a new system for identifying tsunami warning alerts and show the easiest way to go to safety places which have been established by the government.

IV. ANALYSIS

This research expects to develop an application with a map which includes Navigations, GPS, GIS and other related components in order to capture the warning alarm signals from Disaster alarming towers which have been established along the coastal line (Sumanasinghe et al., 2016).

This technology can be used to develop an Android system that captures warning alarm signals from disaster alarm towers that have been established along the coastline. This project uses the map online and indicates the safest places and the fastest route to get there in seconds. The search can make this app available for free download and anyone who travels along the south can be notified if such a disaster warning is posted.

Once the emergency towers issue a warning on an upcoming tsunami, the Android device with this particular application captures the signals transmitted from such a system. Then, the navigation system and the GPS will be automatically activated. Thus, the map could guide the person to the nearest security location on the fastest route. As a consequence of the data analysis, the research revealed that this application should present some of the basic services. Namely;

1. By introducing such an application, it is expected to communicate to people of a probable disaster and thereby, promoting safety and less number of life losses.
2. It enhances to take precautionary measures against and minimize the effect of tsunami.
3. Quick actions can be taken and a wide range of individuals can be informed.

Study area

Figure 1. Sri Lanka Map with Southern

Sri Lanka is an island with an area of 65,000 km which is located between latitudes 5° and 10°N and longitudes 79° and 82°E. The maximum length and the width of

the island measured as 432 km and 224 km respectively. Current population of Sri Lanka is 20,900,516 as of Friday, October 27, 2017 based on latest United Nations estimates.

Basically, this project aims to implement this system for areas in Southern Province. The reason of implementing this project in down south is due to the huge impact Tsunami had on down South coastline in 2004.

A. Data analysis

According to survey which author has done, most of the people are aware of safety places if Tsunami happens again.27 out of 42 in the sample are aware of the safety places. After Tsunami happened in 2004, government implemented several projects featuring with government organizations and non-government organizations. As an example, Helping Hambantota with World Bank, Jaffna Tsunami recovery and rebuilding, Rebuilding Sri Lanka with United Nations etc. Government already established several safety locations in case of Tsunami around Sri Lanka.

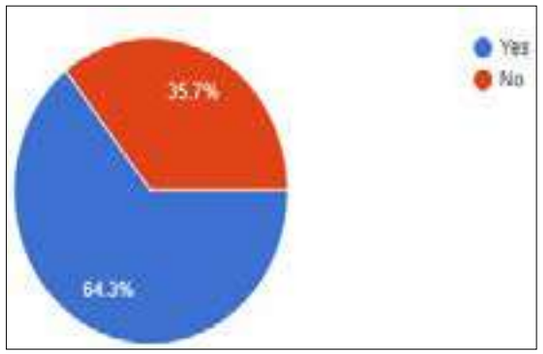


Figure 2. Awareness of safety location in case of Tsunami

The safety locations that have been identifies by this research are, Milidduwa, Unawatuna, Pitiduwa in Galle district, Diyagahagewatta, Hunnadeniya, Kambassawela in Matara district, Malpettawa, Hambantota, Tsunami Helping Hand Housing Project Tissa in Hambantota district.

And also, research has identified the people who contributed to online survey have not taken part in Tsunami pre-training programme.

After 2004 Tsunami, government ordered the Provincial Council to conduct a Tsunami pre-training sessions and government allocated enough financial and non-financial resources for the sessions. Somehow, 88.1% of people who contributed to online survey have not participated in Tsunami pre-training (Papageorgiou et al., 2015).

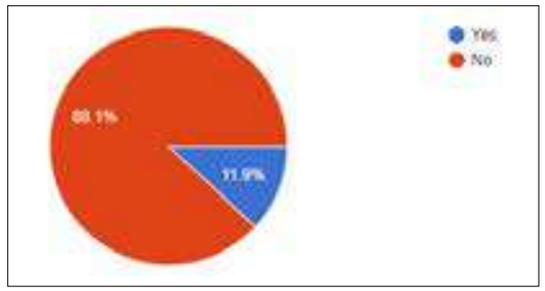


Figure 3. Participation for pre-tsunami training

How is this project related to Information Technology?

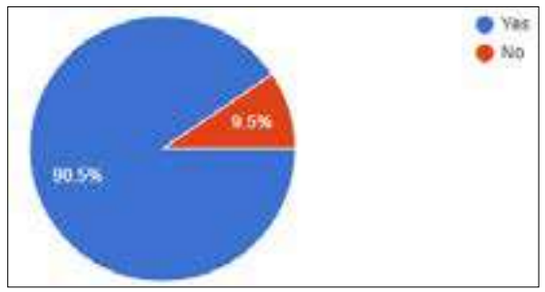


Figure 4. Acceptation of Mobile solution

Once recorded data is analysed, the research should present a solution for problem which arising from the online survey. The problem is, inability to identify obvious way to arrive to a Safety location in case of Tsunami. As a solution, the research suggests an Android application which tracks the Tsunami signals and shows the easiest path or easiest way go to safe location where government has established.

Before December 2004, there was no discussion about the sea disaster. Therefore, there had been no pre-preparation for that kind of situation. The proposed solution given by this research will help people who are living close to costal area and travellers.

V. RESULTS

From the stage of problem identification, information gathering and analysis, this research attempts to solve one of society's burning problems. The main problem that this research will solve is the impossibility of identifying an obvious way to arrive at a place of safety in case of tsunami.

The government has already established the security sites. But to identify a place of safety, a technique must be followed. As the basement of this research, he briefly shows how to perform a GIS analysis to find the safest place in case of tsunami.

In fact, when it comes to finding the safest place in the event of a tsunami, this GIS project has to go through a long process. First, he must create a google form to collect information from the community. Before collecting information, the research must take into account the research that the author intends to carry out. Based on the information collected, factors or parameters are to be collected and prioritize these factors based on the average values.

First, the author must draw a study area. This field returns the project boundaries. After drawing a study area, draw basic layers or basic parameters. According to this project, the parameters are the main roads, the railways, the streams, the streets and the coastal zones. These settings are the shape files of this project. As a beginner, it was difficult when drawing shapefiles for the first time. It takes time to become familiar with the software.

After that, multiple ring buffers should be created for these created shape files. In this too, it was a bit difficult task and after becoming familiar, it was very easy to do the project. After creating several annular buffers, before moving to the union, author had to give the values in the attribute table. There was nothing to think because there were only three attributes in the table. After the union, the layers with the study area, there was nothing but a complication on how to do it (Raju and Sokhi, 2008).

Then, the clipping function was taught. After clipping, by taking the study area as the clipping feature there was no any complication.

Figure 5. Finale map

When it comes to the last steps, author had to merge clipped maps together. In that case, the attribute table was so confusing because there were 30 to 40 attributes in the table. Initially it was felt that it was a hard nut to crack but after learning it was an easy task for author.

In making the final map, there was a confusion because author had 300 to 350 attributes in the table while others had 100 to 200 attributes. After following the steps, the GIS project produces most of the possible areas to locate most security sites in the event of a tsunami.

The final map of the GIS shows all the parameters that have been selected and what the possible areas are to implement the security location. By comparing with the parameters and the possible areas, it is possible to have an idea of what the government must take into consideration before choosing a place of safety. It should be a place that could be reached in the shortest possible time. It must be equipped with infrastructure and have possible means (roads, streets) to reach the security location. Therefore, before selecting the location, the search should consider the parameters and these parameters and the study area are to be totally similar to the basecoat data.

B. Selecting parameters

Selecting a proper land is identified as the number one factor that influences the ability of projecting the business to its full potential. A poor choice of selecting the location is sometimes impossible to re-correct. So that the government must have the ability to select the most suitable location for most safety in case of Tsunami. In selecting a location, many factors are to be considered.

1. Main Roads
2. Railway Roads
3. Brooks
4. Streets
5. Costal Area
6. Junctions
7. Buildings

Identifying each parameter

1. **Main Roads**
   According to this project, government is going to find out the safest location in case of Tsunami. Main roads are the key parameters of this project, due to the easiness to arrive to the safety location through main roads. The identified safety place should be near to main road. If a Tsunami occurs, people can arrive at the location in a quick manner.
2. **Railway Roads**
   Railway roads are another key point of this project. Railway roads are established close to the sea. That is the reason why government has considered the railway roads. The locations which government will be going to find must be far away from the Railway roads.
3. **Brooks**
   There are many brooks located near to study area. The reason why government consider brooks is mere because, brooks increase the impact of the Tsunami disaster. When selecting a safety place, it should be far away from the brooks.
4. **Streets**
   This parameter refers all the lanes and cross streets close to study area. According to this project, one of the main factors is streets. Government has identified these streets are connected to each other. It is like a street network. When selecting a safety location, it should be close to those streets. In case if a Tsunami takes place, people can arrive at the location without wasting time.
5. **Coastal Area**
   According to this project, most interesting factor is coastal area. When selecting a safety location, in case of tsunami, the distance from the coastal should be identified by the government. The safety place should be located at least 1500 meters away from the coastal area.
6. **Junctions**
   Down South area is a kind of very busy area. Near a junction, the traffic is very high. So, the proposed location must be far away from the junction. This parameter is not considered because of low average importance.
7. **Buildings**

   Down South is one of the busiest and crowded areas in Sri Lanka. Therefore, government has a responsibility to find a safety location within this busy area. In this project, junctions are not considered because of low average importance. Buildings decrease power of Tsunami waves.

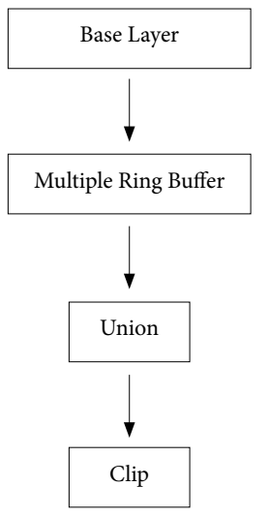
After having selected all the parameters, the priority of the key factors or parameters identified is in progress and checks whether all the parameters have an equal weight or not and then the search can be divided into two categories: Important and Less important (Ware et al., 2000).

After prioritizing the parameters, according to the google form, the data search can select the most important factors selected by the stakeholders via the google form. From these factors, it is necessary to select the most important factors according to the priority, which can be presented in Arc GIS presentation of the map. Author has selected the main factors according to the priority, shown as follows.

1. Main Roads
2. Railway Roads
3. Brooks
4. Streets
5. Costal Area

Figure 6. All the taken parameters

After identifying parameters, project has added Multiple ring buffer, union, clip types of techniques and generate final answer.



VI. DISCUSSION

As a consequence of the development of this research, it benefited people who suffered from the tsunami in 2004 and travellers and people living in coastal areas. The goal of creating this research is to save people's lives if a tsunami recurs. The final production of this research is to be developed as an Android app to track Tsunami alerts and show the most obvious or shortest path to get to the security location.

This research understands that there is no such application or such prevention has been developed previously. There are techniques to track tsunami warnings and also the changes that occur at sea. According to this research, the proposed system is an online application and it should have more specifications compared to other applications.

ACKNOWLEDGEMENT

First and foremost, I would like to express sincere gratitude to my parents who supported with the best to make me who I am today. And also, I convey my gratefulness to Major RMM Pradeep and Mrs. Nirosha Wedasinghe who guided me to do this project and the research.

VII. CONCLUSION

This research is based on guiding people to safe places in the event of a tsunami. Key solution for the problem identified is an Android application to show the most obvious or easiest way to arrive at a safe place and guide people. In accordance with an online survey that involved 53 people, only 64.3% were aware of security sites. The main thing in this research is that it has identified that 88.1% of people have never participated in pre-tsunami training. Therefore, this proposed application will be of immense help for people if Tsunami occurs again.

REFERENCES

Chongyuan, M., Qiang, F.d, n.d. APPLICATION RESEARCH OF TSUNAMI MONITORING AND FORECASTING USING DATA MINING TECHNIQUES.

de Guzman, J.B., de Guzman, R.C.C., Ado, R.G., 2014. Mobile Emergency Response Application Using Geolocation for Command Centers. Int. J. Comput. Commun. Eng. 3, 235–238. doi:10.7763/IJCCE.2014.V3.327

Di Risio, M., Beltrami, G.M., 2014. Algorithms for Automatic, Real-time Tsunami Detection in Wind-wave Measurements: Using Strategies and Practical Aspects. Procedia Eng. 70, 545–554. doi:10.1016/j.proeng.2014.02.060

Juan, T., 2016. Risk Assessment of Computer Network Security in Banks. Int. J. Secur. Its Appl. 10, 1–10. doi:10.14257/ijsia.2016.10.4.01

Olson, M., Liu, A., Faulkner, M., Chandy, K.M., 2011. Rapid detection of rare geospatial events: earthquake warning applications, in: Proceedings of the 5th ACM International Conference on Distributed Event-Based System. ACM, pp. 89–100.

Papageorgiou, A., Tsimi, C., Orfanogiannaki, K., Papadopoulos, G., Sachpazi, M., Lavigne, F., Grancher, D., 2015. Tsunami Questionnaire Survey in Heraklion Test Site, Crete Island, Greece, in: EGU General Assembly Conference Abstracts.

Raju, K., Sokhi, B., 2008. Application of GIS modeling for dengue fever prone area based on socio-cultural and environmental factors–a case study of Delhi city zone. Int Arch Photogramm Remote Sens Spat Inf Sci 37, 165–170.

Romano, M., Onorati, T., Aedo, I., Diaz, P., 2016. Designing Mobile Applications for Emergency Response: Citizens Acting as Human Sensors. Sensors 16, 406. doi:10.3390/s16030406

Sumanasinghe, N., Mikler, A., Tiwari, C., Muthukudage, J., 2016. Geo-statistical dengue risk model using GIS techniques to identify the risk prone areas by linking rainfall and population density factors in Sri Lanka. Ceylon J. Sci. 45, 39. doi:10.4038/cjs.v45i3.7399

Wächter, J., Babeyko, A., Fleischer, J., Häner, R., Hammitzsch, M., Kloth, A., Lendholt, M., 2012. Development of tsunami early warning systems and future challenges. Nat. Hazards Earth Syst. Sci. 12, 1923–1935. doi:10.5194/nhess-12-1923-2012

Ware, R.H., Fulker, D.W., Stein, S.A., Anderson, D.N., Avery, S.K., Clark, R.D., Droegemeier, K.K., Kuettner, J.P., Minster, J.B., Sorooshian, S., 2000. SuomiNet: A real–time national GPS network for atmospheric research and education. Bull. Am. Meteorol. Soc. 81, 677–694.

IMPACT OF ADOPTION OF HOMOMORPHIC ENCRYPTION: SECURITY ENHANCE GUIDELINE FOR SRI LANKAN MILITARY SYSTEM

[1]RMKDLB Abeykoon, RT Udara, GAD Ganepola, WAAM Fernando, RPS Kathriarachchi, and DU Vidanagama

Department of Information Technology, General Sir John Kotelawala Defense University, Sri Lanka

[1]33-its-029@kdu.ac.lk

Abstract- At present, Information security increases the conversation with the occurrences of many data vulnerabilities in current systems. It is now mandatory for all system domains to consider and implement Information security plans. One current procedure which follows for securing information is Data Encryption, especially during end-to-end transmission across computer networks. Encryption is a method by which plaintext or any other type of data is converted from a readable form to an encoded version that can only be decoded by another entity if they have access to a decryption key. Data Encryption has been and still is an area that is continuously being developed. As of today, the latest technology in this area is Homomorphic Encryption - conversion of data into ciphertext that can be analysed and worked with as if it were still in its original form. Ciphertext is plain text exposed to "Cipher" algorithm which is applied to plain text to get ciphertext. The authors present here the applicability of this technology on Sri Lankan Military Domain. The methodology used to conduct this research is a qualitative and quantitative based survey. The online survey was circulated through e-mail and the survey was successfully completed. According to the survey it could be analysed that the security when transferring data/information in this domain is very low-grade, which in contrast, must be very high due to the presence of sensitive data related to national security of the country. The authors have designed in detail on a set of recommended guidelines for secure transmission of military data using this technology.

Keywords- Information Security, Homomorphic encryption, Encryption Technology, Adoption of encryption.

I. INTRODUCTION

In computing, Encryption is the methodology by which plaintext or any other type of data is converted from a readable form to an encoded version that can only be decoded by another entity if, they have access to a decryption key. Encryption is one of the most important methods for providing data security, especially for end-to-end protection of data transmitted across networks. There exist many modern and emerging technologies of Encryption such as Searchable Encryption, Homomorphic Encryption, Functional Encryption, etc.

Homomorphic Encryption in simple definition is, conversion of data into an encrypted format (i.e. ciphertext) and allows operations, functions, etc. to be performed on that format without the use of decrypting. Considering a real-world scenario, very important documents are kept inside a safe. However, it has to be taken out to perform any work in the documents. This creates a vulnerability of a certain data breach to occur. The above vulnerability can be discarded if work in documents can be done inside the safe. Homomorphic Encryption follows the exact scenario above. This technology allows complex mathematical operations to be performed on encrypted data without compromising the encryption. For instance,

imagine two numbers that are not encrypted: 30 and 50. Once encrypted, the numbers become 43 and 72. The sum of the two encrypted numbers results in 115 (43 + 72). In homomorphic Encryption decryption of 115 the results in 80 (which is equivalent to the sum of the original numbers, 30 + 50).

The above technology of Encryption is used in many application domains where there is a high requirement of complex mathematical calculations which needs to be performed under high security for instance, high-end calculations in space stations, nuclear power plants, military operations, Bit coin mining, etc. In fact, Homomorphic encryption is an advantageous technology to avoid/minimize data breeches when data is being processed. This yet-developing technology attracts many researches to be performed. In this research paper, the authors researched on the application of Homomorphic Encryption into the operations of Sri Lankan Military. They have conducted a detailed requirement analysis by interviewing various personnel of the tri-forces and have recommended a guideline of methods of the application of Homomorphic Encryption.

II. OBJECTIVE OF STUDY

The authors scrutinize feasibility of applying homomorphic Sri Lankan Military domain with the aim of providing a higher data security during Military communication.

III. OVERVIEW OF METHODOLOGY

The authors formed a hypothesis which was tested by the responses gathered via an online survey. The authors then reviewed the responses and checked the correspondence between the responses and their hypothesis developed earlier in the process, thereby proposing a practical guideline of a real-time application.

The online survey was carried using a set of Military personnel in the Signal's unit of Sri Lankan Military domain.

IV. DATA ENCRYPTION: DEFINITION & PROCESS

Simply defining, Data Encryption is converting a message into another type of format so that the message meaning is not obvious.

To perform the encryption, it is essential for an encryption algorithm and an encryption key. The former being the mathematical calculation for the conversion and latter being the unique string of bits that determines the transformation through the algorithm.

The process of Encryption is as follows. Unencrypted data, plaintext, is encrypted using both the algorithm and the key. This process generates ciphertext that can only be viewed in its original form if decrypted with the correct key. Decryption is the reverse of encryption, the same steps but inversing the order in which the keys are applied.

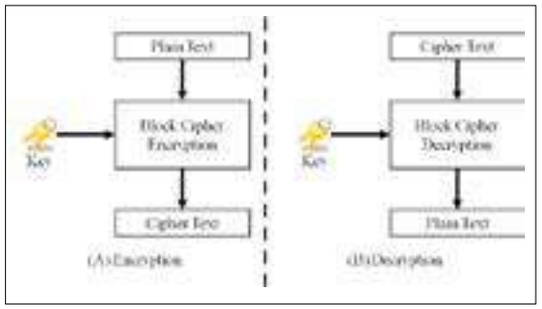


Figure 1. Diagram of encryption and decryption in Block Cipher
Source: Research Gate

Presently used encryption algorithms fall into two categories: Symmetric and Asymmetric Symmetric algorithm: Symmetric-key ciphers, also referred to as "secret key," use a single key. The most widely used symmetric-key cipher is the Advanced Encryption Standard (AES), which was designed to protect government classified information. Symmetric-key encryption is usually much faster than asymmetric encryption, however, the sender must exchange the key used to encrypt the data with the recipient before the recipient can perform decryption on the ciphertext. The need to securely distribute and manage large numbers of keys means most cryptographic processes use a symmetric algorithm to efficiently encrypt data but use an asymmetric algorithm to securely exchange the secret key.

Asymmetric algorithm: Asymmetric cryptography, also known as public key cryptography, uses two different but mathematically linked keys, one public and one private. The public key can be shared with everyone, whereas the private key must be kept secret. The RSA encryption algorithm is the most widely used public key algorithm,

partly because both the public and the private keys can encrypt a message; the opposite key from the one used to encrypt a message is used to decrypt it. This attribute provides a method of assuring not only confidentiality, but also the integrity, authenticity and non- reputability of electronic communication and data at rest using digital signatures.

V. HOMOMORPHIC ENCRYPTION: AN INTRODUCTION

Homomorphism as described in the research paper of Ogburn et al. "a mapping of a mathematical set into or onto another set or itself in such a way that the result obtained by applying those corresponding operations to elements of the first set is mapped onto the result obtained by applying those corresponding operations to their respective images in the second set."

The process of how this encryption methodology works has been explained clearly by Brian Hayes in American Scientist Journal via an article named "Alice and Bob in cipher space"

Consider two sets of data with one set, of positive real numbers, R, and the other, logarithms of real numbers. The multiplication of real numbers and the addition of logarithms are considered to be homomorphic operations. Considering x, y, z to be real numbers,

First operation: x *y= z

Second Operation: log(x)+ log(y)= log(z).

There are two ways of obtaining the result z. First is to apply the operations to t plain text/unencrypted data and the second to encrypt the data (in this case takin the log values) apply a different operation to get a result and then converting the result to get the intended result. (taking the antilogarithm of z gives the result z).

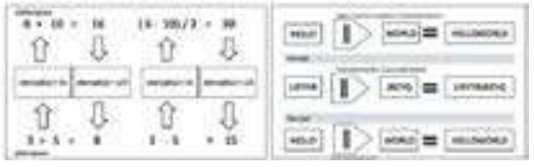


Figure 2. Encryption and decryption example
Source: American Scientist

V. COMPARISON OF HOMOMORPHIC ENCRYPTION WITH OTHER EXISITNG ENCRYPTION STRATEGIES

Figure 3. Basic message sending mechanism
Source: Author

Figure 3 depicts the traditional process of information transmission. Security and Encryption for any messages is unavailable which in turn allows any 3rd party to access and utilize the information.

Figure 4. General message sending mechanism with data encryption
Source: Author

Figure 4 depicts the basic encryption strategy most parties use presently. The messages are encrypted into symbols and characters which are then transferred over the Internet or Intranet-work. However, anyone with the access to the decrypt key of the encryption strategy will be able to decrypt and access information.

Figure 5. Message sending mechanism with homomorphic encryption
Source: Author

depicts the stages if the homomorphic strategy. Information is first homomorphically encrypted giving an output of with readable phrases but of complete irrelevance to original information's meaning. Next the homomorphic encrypted information is encrypted using normal strategies transforming to symbols and characters and transferred to the destination node via untrusted media.

The reason for encapsulating the homomorphic encrypted information using normal encryption strategies is to deceive the intruder that in case he captured the information and decrypted, he will be misled by the decrypted information because the decrypt information is provides meaningful phrases but is of complete irrelevance to original data

VI. IMPORTANCE OF DATA SECURITY IN MILITARY COMMUNICATION

Communication of data occurs through Wired Computer networks, Wireless networks or through physical means, that is man-to-man transmission. With the advancements of technology over the 21st Century, transmission of via physical means have reduced due to inefficiency & poor data security when compared to Wired/Wireless Computer networks. Data Security whilst transmitting through this mean is important due to the nature and quality of data being transmitted. Sri Lankan Military system uses an Intra-network and microwave based wireless technology for their communication. These networks constantly transfer to and from data that is vital for national defense of the country. It is mandatory that the data being transmitted need to be protect its confidentiality and kept out of reach of those any who are attempting to disrupt the nation's security. Hence, Information and Data Security acts as the "Heart of Military communication."

According to analysed data from online survey, the particular domain does not incorporate any such effective data encryption while performing data transmission. Since this system consist of highly confidential data related to national defence it is mandatory for potent data encryption strategy.

VII. APPLICATION OF HOMOMORPHIC ENCRYPTION TO SRI LANKAN MILITARY SYSTEM: A RECOMMENDATION

When compared to other Encryption technologies, this provides high end security for point-to-point transmission of military data. A scenario of the application would be, consider an important document regarding the current status of national security is sent from Defense Headquarters to a remote Military base. The document would be encrypted homomorphically. An algorithm would be conversion of letters in the document into another set of letters which would give another meaningful word but completely irrelevant to the originally transmitted document.

For instance, consider an original message from Colombo to Hambantota military base is sent. The message may include "Announcement to all Officers". When it is being transmitted over the intra-network it will be transmitted as "Roll Call Scheduled at 0800HRS". An intruder who intercepts the transmission media from middle are false guided by the homomorphic encrypted message. The intruder may be deceived of the transmitting messages as to not having encrypted, because the standard encryption implementations are producing ciphertexts which is a collection of letters, strings, symbols and tend to be encrypted.

Once the message reaches destination, it decrypts the data to the original message using homomorphic algorithms.

VIII. METHODOLOGY OF STUDY

A. Hypothesis

Based on Background Study, the authors developed a hypothesis which state that:

Homomorphic encryption is the optimum encryption strategy for the use of Military Communication due to the fact that:

a. During war & other crisis, military communication within their and among other teams play a key role. Locations, mission's statuses, war plans, daily objectives, etc. are some sensitive information being shared. Therefore, it is very much important to have a proper way to secure those messages when transferring from one place to another.

b. Current data encryption method or the mechanism Sri Lankan Military uses does not have a method to verify or retrieve all the messages on demand and it's a complex task to maintain a proper code (key) for every message. To send a message, a code is created particularly for the specific content and establish a connection across those two communication centers. At the receiving center, the location is confirmed message is retrieved using that specific code. Once the communication is over, the code that was used will not be used any longer, and a new code will be generated instead. But those messages are relatively very short messages which in turn makes it highly prone towards the risk of attackers and hackers.

c. With the introduction of homomorphic data encryption method, Sri Lankan Military can share their information about Locations, Mission's statuses, and other relatable information. Since this technology is yet under the development and in the terms of emerging technology, encryption is only possible for numerical information, such as locations, GPS status, and any other forms of communication which is done by numerical terms.

d. Main concern for the use of homomorphic encryption is because when the message is encrypted homomorphically, calculations can be performed while keeping the message in the encrypted state. Therefore, anything included in the message will not be decrypted, and hackers and attacker will not be encouraged to break such mechanisms very easily and read what was encrypted. Even if the calculations are done without decrypting. Final outcomes will be accurate and perfect. For an example if X=4, Y=5 and Z=1. when we decrypt it. Values of X, Y and Z will not be visible, but according to the homomorphic method we can still do calculations and get the accurate answer without decrypting those values.

e. Therefore, Homomorphic Encryption will be very much useful for Sri Lankan Military to share their information and statuses securely and in a most efficient way, and attacker or hackers will not be able to decrypt the message or read it.

B. Data Elicitation

The online survey was conducted with the intention of gathering responses to the following questions:

a. In the perspective of the militaria, Does Sri Lankan Military domain require to deploy any other encryption strategy apart from the current that is being used?

b. Does Sri Lankan Military domain be operationally feasible to adopt an emerging encryption strategy?

c. If, Homomorphic encryption was introduced as an emerging encryption strategy into the domain, would this be able to achieve the requirements of a secure military communication (in the context of militaria)

A group of interviewees of higher ranked officers in Sri Lankan Military Signals Unit were selected. The authors invited 120 personnel from which they received 100 responses. An online questionnaire series developed from Google Forms were sent to each personnel.

The authors selection behind the following interviewees is that:

The following questions were presented to the interviewees:

Q1. How is your knowledge on computer-based software and other IT related technical skills?

Q2. Are you aware that data security is now a prominent role in IT infrastructures?

Q3. In your opinion, do you consider that data security implemented at present meets with its current functionalities?

Q4. How would you recognize the reliability of the chain of command inside your organization?

Q5. Your opinion on the security of physical data transferring method (man-to-man)?

Q6. Your point of view on usefulness for updating of current data securing techniques?

Q7. Your opinion on your knowledge on data encryption?

IX. RESULTS & DISCUSSION

A. How is your knowledge on computer-based software and other IT related technical skills?

Majority of the Military personnel who were surveyed had an average knowledge on computer-based software programs and other IT skillset. As per the weights, the authors received an Average value of 80%, 10% of good and 7% 2% 1% on Very good, low and very low respectively from the Questionnaire being provided.

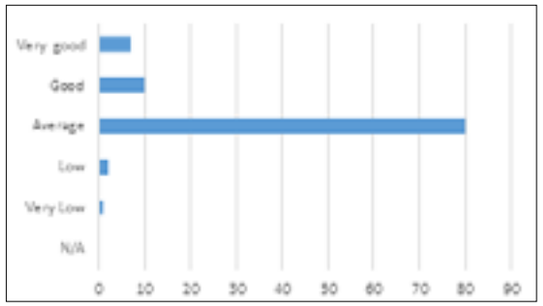
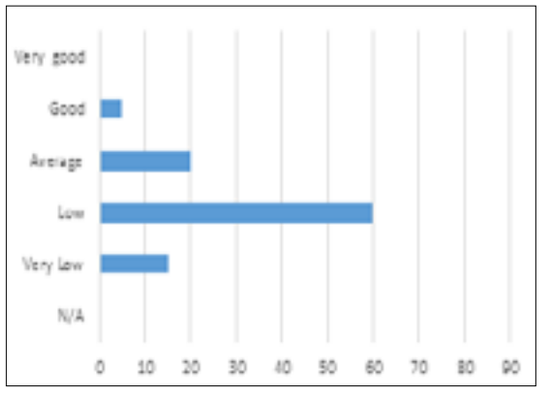


Figure 6. Analysed Data Chart
Source: Author

Based on the above figures, the authors observe that majority are proficient on the fast-changing Information Technology. With this base on mind, the authors drew a conclusion that majority personnel in the tri-forces would comprehend Homomorphic Data Encryption and accept & support for the implementation of this new technology in their domain system.

B. Are you aware that data security is now a prominent role in IT infrastructures?



Source: Author

The majority who answered this were those having sound knowledge on IT and working in the field related to IT. However, the weights showed a poor level on knowledge regarding data security. Weights were as follows 60% low, 20% average, 15% very low and 5% good. The authors were clear that there is a serious need of educating the personnel on importance of data security and enhance them with the value of Homomorphic data Encryption method.

C. In your opinion, do you consider that data security implemented at present meets with its current functionalities?

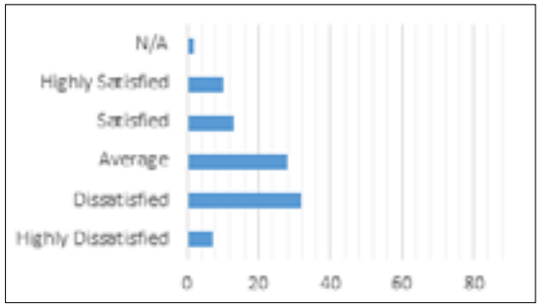


Figure 8. Analysed Data Chart
Source: Author

The authors revealed that wireless data transmission technology used currently is via microwaves. There is null deployment of data encryption technologies, which means Sri Lankan Military communicate over untrusted networks using plain texts.

Analysed data provide on the knowledge they have on their current data security and its functionalities, 32% of the personnel aren't happy with current data security functions and 28% of personnel are on a dilemma if the current data are well secured or not, therefore authors suggest data encryption and data security countermeasures must be applied in order to overcome vulnerabilities of current data security.

D. How would you recognize the reliability of the chain of command inside your organization?

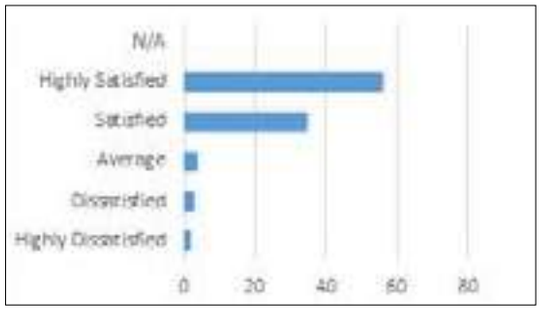


Figure 9. Analysed Data Chart
Source: Author

According to analysed data, 90% of personnel are on high recommendation of the reliability of their chain of command, therefore authors suggest the usage of homomorphically encrypted data inside the domain system as a method of securing data before any transmission occur for strictly permitting access for higher ranking officers.

E. Your opinion on the security of physical data transferring method (man-to-man)?

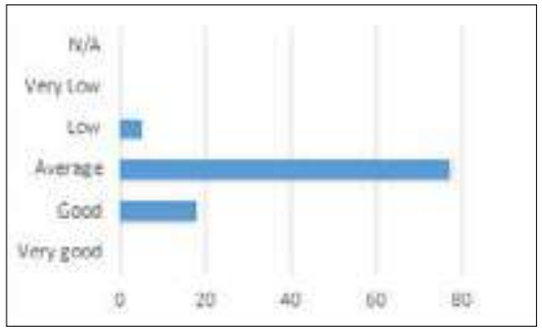


Figure 10. Analysed Data Chart
Source: Author

Most personnel consider that the security level is average in man-to-man data transferring technique. As shown in the results we got 77% of Average, 18% of good and 5% on low and 0% for the very low and very good from the survey. With the analyzation of these data authors concluded that there is an evident need of more secure method in physical data transferring.

F. Your point of view on usefulness for updating of current data securing techniques?

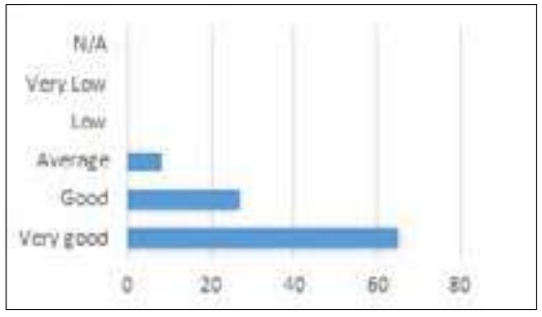


Figure 11. Analysed Data Chart
Source: Author

Majority's perspective was to use more updated data security techniques. As shown in the results weights obtained were got 65% of very good, 27% of good and 8% on Average and 0% for the low and very low. On analysation, authors concluded that there is an urge of more secure and updated method of data transferring such as Homomorphic data Encryption is required.

G. Your opinion on your knowledge on data encryption?

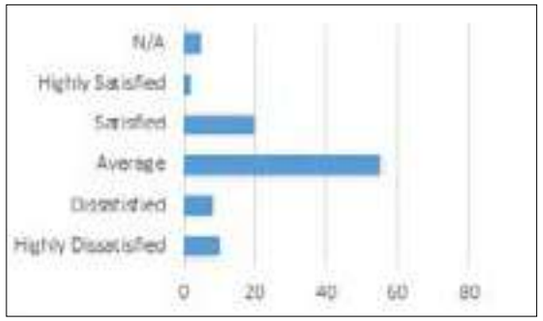


Figure 12. Analysed Data Chart
Source: Author

55% of the personnel who faced the survey are self-confident on their knowledge on data encryption. However, with consideration of every suggested guideline that is to be applied, they may have poor knowledge on data encryption and data security, based on as collected data, personnel are not familiar with data encryption and data security. Therefore, knowledge about data security and data encryption must be increased to update their data transmission techniques and current data security countermeasures.

X. CONCLUSION

Based on the above data gathered by the authors through online questionnaire, it is evident that majority of military personnel are aware of the theoretical background of Data security and its methodologies and its significance. However, practical implementation of Data Security strategies is not being implemented/ are being implemented but in very poor manner.

The authors came into an implication that it is a drive force that is required to integrate the domain system with

the practical implementation of data security strategies. Hence, it is feasible to deploy an emerging encryption strategy like Homomorphic encryption in the domain.

Based on information gathered from military personnel, the authors' hypothesis was proven to be correct.

XI. FURTHER WORK

This paper introduces Homomorphic Encryption to Sri Lankan Military System to secure the Confidentiality, Integrity and Availability of data that is being communicated over transmission media. The authors have presented a recommendation of application of a Homomorphic encryption algorithm to minimize attacks.

Further works of this research include in developing a mathematically correct algorithm that would achieve the output of the designed algorithm, also can be discovered relationships and core relationships by doing this research more advanced using STSS software.

REFERENCES

B. Hayes, "Alice and Bob in cipherspace", American Scientist volume 100, 5,2012

Ogburn M., Turner C., Dahal P. , "Homomorphic Encryption", Procedia Computer Science(20), 502-509,2013

Pfleeger, C. and Pfleeger, S. (2012). Security in computing. Upper Saddle River: Prentice Hall.

Definition of Ciphertext | What Is Ciphertext ? Ciphertext Meaning. - The Economic Times. Accessed June 05, 2018. <https://economictimes.indiatimes.com/definition/ciphertext.>

Decrypting Data. - Accessed June 05, 2018. <https://msdn.microsoft.com/ja-jp/library/te15te69(v=vs.85).aspx.>

What Is Encryption? - Definition from WhatIs.Com. -SearchSecurity. Accessed June 07, 2018. <https://searchsecurity.techtarget.com/definition/encryption.>

Definition of Decryption | What Is Decryption ? Decryption Meaning. - The Economic Times. Accessed June 07, 2018. <https://economictimes.indiatimes.com/definition/decryption>

AN EFFICIENT WEB ENABLED AUTOMATIC EMERGENCY MEDICAL ASSISTANCE SYSTEM USING ANDROID

UDVA De Silva[1], HRWP Gunathilaka, and N Wedasinghe

Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

[1]*kdu-32-it0010@kdu.ac.lk*

Abstract- This paper introduces a system to manage medical assisted emergencies in certain areas. At present, Sri Lanka and many other countries in the world face many difficulties in managing emergency medical situations even though telecommunication has been developed and it provides the technology needed for this kind of healthcare systems. Scarcity of Infrastructure and resources needed for such situations is a major problem in many areas of the country and in the world. The probable cause for this is that the relevant authorities do not pay much attention to this problem. As these situations occur abruptly without any warning, the implementation should be mobile. Thus, the mobile part has been developed on android platform. It will automatically send the location to the guardian and the same app will collect the information and the, location of the patient, and it will inform about the problem to the health care centers or the emergency services in the relevant areas. Through this system, public can directly connect to them, and they can get a fast and efficient response from the emergency medical assistance team who will receive patient's information from the web application.

Keywords- Location, GPS, Android, Emergency Medical Assistance, Web-based patient information

I. INTRODUCTION

Despite being a developing country in the 21st century, Sri Lanka faces many health issues, mainly due to the busy life style of people. Among these issues, there can be emergencies where medical assistance is needed immediately. However, unfortunately there are not enough medical assistance to cater the so-called problems in the country due to several reasons, such as the low budget allocation on medical services and public awareness in protecting the patients. At present the common method used to manage those kinds of situations is to give an emergency contact number to the public so that the public can contact the authorities in such situations. It is not a sufficient way to deal with emergencies because so many errors and delays can happen, and it costs lives. That is the main problem identified. These methods are not considered as user friendly as they create serious problems. The hypothesis is that the emergency medical assistance system in Sri Lanka can be improved and it will revolutionize the medical assistance sector in Sri Lanka.

The Information & communication technology can be used for this purpose because right now Sri Lanka does not have a proper ICT based emergency medical assistance system. Members, admins, police, medical officers, and developers can access the emergency medical assistance system I present in this paper. There are various levels of access for users, but the admin only can control the server computer because he only has the admin right to the system.

The "Suraksha" mobile application allows the public to inform the authorities easily and quickly about the locations of the emergency situations.

Anyone can access this mobile application from anywhere by typing their user name and their password. And then

he/she can track the emergency and send it to the nearest health care centre or emergency service. Then user will get a confirmation message whether the tracked location and other information (date and time, username, user mobile number and health related information of the user) are received or not by the relevant health care centre or emergency service.

A. Aim

Aim of this application is to facilitate people who need emergency assistance by providing emergency services and help to those who require a certain need or have a disease and to make their lives easier.

B. Objectives

According to the aim of developing an Emergency Medical Assistance System, the following objectives were identified.

- Analyze the problem and understand the problem clearly.
- Study and identify the technology used before.
- Transforming the requirements into a form implementable using a programming language.
- Conversion of the design into a working code.
- Identify the existence of the faults in the new system.
- Maintain the system after delivering to the user.
- After submitting the proposed system do a research to find out further improvements.

C. Hypothesis

The Emergency medical assistance in Sri Lanka can be improved by developing a mobile-based automated system and it will revolutionize the medical assistance sector in Sri Lanka.

D. Software solution for Emergency Medical Assistance System

The solution proposed in this paper is to develop a mobile-based emergency system to medical related situations which will be more suitable for developing countries. The introduction of mobile phones in this process will enhance productivity. The system will be developed in two separate components; one part is a mobile-based application and other part is a web-based system that will work as a server for this system. According to the requirements, web application has been developed by using C# and SQL database to run on windows operating system. The mobile application has been developed by using Java, Android studio and MySQL. The mobile application registration part developed by PHP.

E. Resource Requirements

- The mobile application will be compatible with any mobile that runs on android.
- The web server will be developed to be compatible with any computer that runs on Windows operating system.
- The mobile application will be implemented by using Java and XML in android studio.
- The web application will be developed by using the C# language with ASP.net framework in visual studio.
- Web application database will be implemented by using DB forge studio in SQL server.
- The MySQL server is also used to work as a mobile application database server.
- The login and registration of mobile application server side is handled by PHP language.

II. LITERATURE REVIEW

The review will be focusing on the procedure and available practices in previous Emergency Medical Assistance Systems.

Emergency Services Using GPS Navigation(Sarkar, 2016); the main aim of this project was to make sure the patient reaches the hospital as soon as possible. There are two main systems integrated in this system to become a functional system. They are Arduino based GPS module system and android based software solution. In this system the GPS module is planted on the ambulance so that the hospital management can keep track of the ambulance and in case of emergency they can directly send the ambulance to the patient's location.

The Emergency Medical Help System Using OSM (Burkul et al., 2015) project has android based mobile system developed for detecting nearest hospitals for the user. The system has a database as a central server. It has the detailed information about the hospitals. The android application is installed in the user mobile device. The application will show the nearest health care centers to the user using open street map. It has a special functionality for the visually impaired people. it will provide directional audio information of the location to the patient.

Most of the emergency alert systems are installed on android or any other type of system device and wired or wirelessly linked to a gateway and then connected to a hospital. Most of these systems are divided into two systems as mobile and embedded, but this is not an effective way to build a system, because running two systems in same time is a wastage of resources. Therefore, the best strategy is to install the system on a mobile device. (Jadhav et al., 2016). The emergency alert system must be clicked manually or automatically. The alarm then sends the emergency message to all registered patient families and the doctors and with that info, it will carry the location information of the user using GPS system.

In India, the number of road accidents has rushed in a very drastic manner. In addition, after a further observation, the researchers understood that this problem is very much severe in modern urban areas. In this paper they proposed an automatic ambulance management system (Anand and Flora, 2014) and this system is reinforced by the intelligent transport system which would control the traffic light systems. So, the system will detect the location of the accident occurred, and then the system sends the nearest ambulance to the situation. When travelling to the incident the system will control the color lights and make path to the ambulance, and after the ambulance gets the patient, the system will clear the path to the nearest hospital for the ambulance by changing color lights. This will reduce the time significantly when getting the patient to the hospital.

There are many tracking systems but most of them are expensive, so these developers tried to develop a Cost-Effective GPS-GPRS Based Object Tracking System (Hasan et al., 2009) which will record the positions of the object using GPS and will show the location using an embedded google map. In this system, the SMS technology is used to communicate with other devices.

The previous researchers have done a tremendous amount of work to develop a proper emergency medical assistance system. They done a good job too. However, the development of technology has made those systems out dated, much complicated, and very expensive to implement and maintain. Therefore, this proposed system is a timely requirement for health care sector.

III. METHODOLOGY AND EXPERIMENTAL DESIGN

A. Survey

Before developing the system, the authors decided to do a survey to get a brief idea about the user requirements, previous systems, popular mobile OS, suggestions for new system and so on. After doing this survey (81 responses), the authors developed the initial planning for the new system.

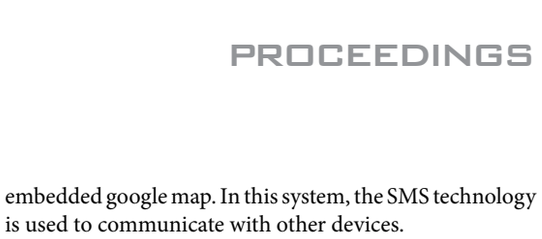


Figure 1. previous Medical assistance systems developed in Sri Lanka

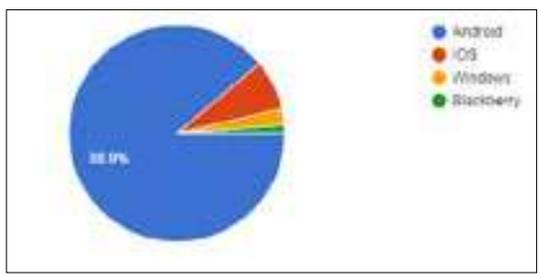


Figure 2. The favourite mwobile OS type

Figure 3. Suggestions for Medical assistance System

B. Data Gathering

Qualitative and quantitative data required for designing the requirement specification for the Emergency medical assistance system was gathered through observation, surveys and document review techniques that were carried out with various types of people around Sri Lanka. Face to face interviews with lecturers and students were also conducted in the process. A survey was conducted by distributing a questionnaire among students, lecturers and public, and then the answers were gathered accordingly.

C. Data Analysis

The data which were gathered during the data gathering phase was analyzed by using various analyzing techniques. By analyzing the data, the group could reveal the extent procedure, problems, limitations, suggestions of the students and lecturers for the emergency medical assistance system. The drawbacks that were identified in the existing solutions are: not sufficient server systems to handle user requests, only the mobile device shows the way to medical care center and it is a very risky approach to handle patients; using outdated technologies like SMS to send the location of the hospital (its address) as a message; some systems are developed to handle only one type of emergencies like vehicle accidents; some systems are not completely developed and so on. Concluding the analysis section, the authors finally decided to design an emergency medical assistance system with the ability to alleviate all the above- mentioned drawbacks.

D. Approach

In the Emergency medical assistance system process, first the user should login to the Mobile application by providing user's name and phone number to the system. Then the mobile application will track the coordinates (longitude and latitude) of the location where the user is in (where the emergency occurs). Then that information: user name, user mobile number, date and time, coordinates and health related information of the tracked emergency will be sent to the web server which is in the health care centre or emergency service. After taking the information from the mobile application those details will be displayed on the emergency service web server. The location of the emergency will be displayed on a map according to the coordinates that were sent from the mobile application. Then the assigned people will be sent to those places to collect or treat the patients. The web application will send a notification to the user whether the information was sent successfully or not and will give a compliment for the sender for his service. Each information that the users send will be stored in SQL server database successfully. A continuous internet connection is required to carry out the functions of the application.

E. Technology Adopted

It is very important to use acceptable tools to develop a productive system. Use of any inappropriate tools can solely affect the developed system with unnecessary errors and faults and use of those erroneously chosen technologies additionally can cause the new system to be crashed on the implementation stage. Correctly chosen technologies will enable manufacturing a system with top quality. These technologies may result in developing a system that do not spend lots of time and resources to perform a task that is anticipated by the system. It is very important to use applicable programming language and the other necessary tools to develop a productive system. Further, these technologies and tools can help to develop the system within a minimum development time. The most important objective of developing this type of an application is to produce a more efficient working system, instead of using manual approach. Hence, we should use the most applicable tools available in the market to develop the system. Technological considerations followed during the development of the system are efficiency and performance, re-usability and flexibility, and object-oriented development support. So according to the Emergency Medical Assistance System, java, android studio and MySQL server are used to develop the mobile application module. The mobile application server configuration is done by PHP and is hosted by xampp server. C# language and SQL Server 2014 are used to develop the database for this project. The web application is hosted on IIS server. This chapter includes the details about the technologies that we are going to use to develop Emergency Medical Assistance System.

F. Application server (web)

The programming language that is going to be applied as the main developing language for the system development will turn into significantly trusted accuracy and performance. When considering all the technologies that can be associated with the proposed system, it can be identified as a web-based technology.

The .NET Framework contains the common language runtime and the .NET Framework class library. This class library is a complete, object-oriented collection of reusable types that you can use to develop applications like traditional command-line or even graphical user interface (GUI) applications, to applications based on the latest innovations provided by ASP.NET (A child of a .Net frame work), such as Web Forms and XML Web services. It is a very powerful framework and a solid platform to develop web-applications.

The CLR is the implementation by Microsoft for the common language infrastructure (CLI). It helps to create execution and development environments in which languages, and libraries work together without any flaw. CLR provides other services such as automatic garbage collection, exception handling, and resource management.

Those concepts used to incorporate with the emergency medical assistance system for better functionality with web services is a key feature in this solution.

G. Mobile Application

Android gives you the freedom to put into effect your own device specifications and drivers. The hardware abstraction layer (Hal) presents a well-known approach for developing software hooks among the android platform stack and your hardware. The android working machine is likewise open source, so you can contribute your own interfaces and improvements. Android is an open source, Linux-based software stack created for a wide array of devices and form factors.

GPS location fall detection using acetometer and using two web servers are novel concepts used to incorporate with the emergency medical assistance system for better functionality and it is the key feature in this solution. In android the developer can freely access those functionalities easily because of the open source license.

H. Database Selection

Consistent with the above eventualities most of the structures use the square database to keep facts because it is simple to control and perform. Therefore, the database that put in force on the server must be able to supply efficient operations. Consequently, the proposed emergency medical assistance system is using the Microsoft SQL server 2014 as server. SQL server is the inspiration of Microsoft's data base platform, delivering challenge critical performance with in-reminiscence technology and quicker insights on any information, whether on-premises or in the cloud, and Microsoft SQL Server is an application used to create computer databases for the Microsoft Windows family of server operating systems. Microsoft SQL Server provides an environment used to produce databases that can be accessed from workstations, the Internet, or other media too. Database management or DBMS, stores user's data and enables the information to be transformed into statistics. However, the mobile application login process will be handled by the MySQL database. It is a very light weighted database system that can call upon request from user. It is very easy to use with java and PHP languages that are used to develop mobile application, so it is a vital functionality for the system.

I. Software Design

Programming the Android Application: For the programming of the Android App, the Android Studio 2.1.1 is used. It is a well-recognized developing platform for android development.

Programming the Web Application: For the programming of the Web application, the Visual studio 2015 is used. It is the Microsoft recommended platform to develop web applications in windows. When it comes to development,

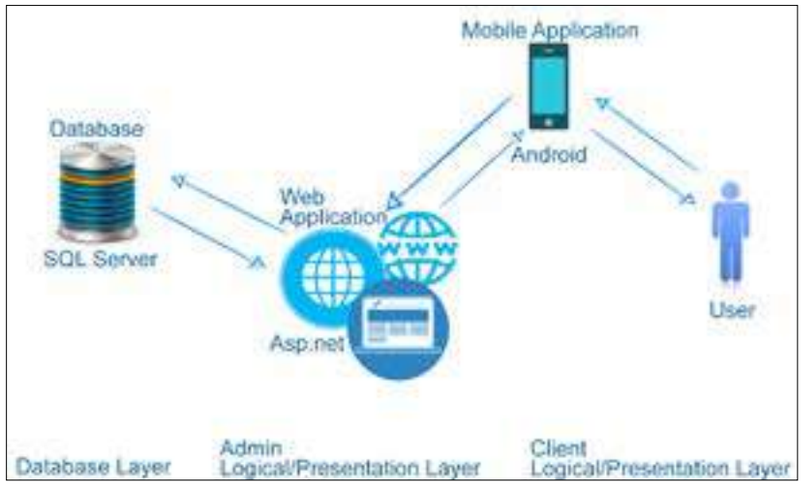


Figure 4. Overview of the proposed system

it is an unmarked territory to the developer, so the resource materials are vital for the visual studio. There is a good availability of resource materials.

To develop the mobile application, server side is developed by PHP scripting language; most of the coding is done using the sublime text editor because it is good in suggesting required words in required time.

IV. THE PROPOSED SYSTEM

Figure 4 shows the architecture of the proposed system. The system is based on a smartphone application that continuously detects whether there is an emergency, using the built-in accelerometer. In the case of an accident, the location is identified using the built-in Global Positioning System (GPS). The system then sends a message to emergency services and to the registered emergency contacts notifying them of the user's information, accident, and its location.

The system consists of two main components: an android

application and an application server. Each is described below.

A. Mobile Application

The mobile application, as shown in Figure 5, is with a built-in accelerometer and supporting smartphone

Figure 5. Screen shots of the mobile application.

location services like built-in GPS / GSM triangulation and fall detection. In the registration, the application allows a user to enter their personal information including name, ID, blood type, and phone numbers of individuals to inform in case of an accident. If an accident occurs, the application immediately sends an SMS to the guardian and emergency services, with user's information and geo-location. Before sending SMS, the application alerts the user about the fall-detected alert on smartphone screen. The application also gives the user an option to stop a false alarm.

B. Web Application

As shown in Figure 6, the admin web application is built using IIS, C#, and SQL server. The application server provides the following services.

- Real-time reporting of emergency with geolocation
- Various reports showing current emergency and their locations.
- Showing the patient information that is needed to treat him or her.

The login password is encrypted using the SHA-1 (Hash functions are intended to achieve a "one-way transformation" the original message cannot be recovered from the digest, at all.). There are two choices for the admin: either he or she can access the information that is sent by the user or he can directly access the map that has the exact locations of the patients; or the information sent by the user are displayed in a table. The last raw of the information table contains the location link. By clicking that, the admin can access the location of the user in the map.

Figure 6. Screen shots of the web-application application

The mobile web application is a web-based application without an interface-built. It functions using Apache, PHP, and MySQL server. The application server provides the following services.

- Real-time user registration.
- Saves the user information to the server.
- Retrieves information from the server when needed.

V. EXPERIMENTAL RESULTS

The fall detection was executed on each of the test cases to make a prediction about the actual state. The results are shown in the form of a confusion matrix in Table 1. The confusion matrix shows the comparison of predicted vs. actual results. For example, in Table 1, out of 50 no-accident cases, 46 were predicted correctly by the application and only 4 cases were false positives. Furthermore, all 150-accident cases were predicted correctly as accident cases.

As the Table shows, the fall detection module was able to achieve an overall performance of (46+150)/100 = 98% accuracy in distinguishing between a no emergency and an emergency state.

Table 1. Confusion Matrix for System Testing

Actual / Predict	No emergency	emergency
No emergency	46	0
emergency	4	150

VI. EVALUATION

Summative evaluation was used as the evaluation method to find how the system functions and whether it is up to the expected level to fulfil the users' requirements.

The intended target group of the emergency medical assistance was not employed for testing purposes as a safety measure, since the objective of this project is to design and construct a prototype emergency medical assistance. Therefore, its functionality was tested with the help of colleagues who supported in the testing and evaluation by giving their opinions after using the emergency medical assistance, module by module.

Several colleagues were given mobile phones with the installed mobile application and they were connected to a same Wi-Fi router so the distance from one user to another is low. When the user initiated the application, it worked very accurately and faster than expected. Sometimes the users sent requests at the same time and it worked fine even on that. Some coordinates of other countries were given for test purpose. Even in global scale, the system worked fine. According to the colleagues who used the system, they were very satisfied with it.

As this system focused on designing and constructing a prototype of emergency medical assistance, comparatively cheap Wi-Fi router was used with a fixed IP address. When

implementing the actual medical assistance system, it is recommended to use public IP with a higher accuracy to minimize errors in sending messages to the correct server.

VII. CONCLUSION

The author's ambition was to develop a cost-effective emergency medical assistance system for public. The developer implemented this system to determine its ability to satisfy the entire functional and non-functional requirements with special qualities such as flexibility, reliability and efficiency. Another aim was to overcome the drawbacks identified in the existing systems and new system.

There is a mobile application and a web-enabled system, so this mobile application offers the user to install the application and enter data. In the Web-application, admin monitoring records are sent earlier. Data storage and retrieval will become faster and easier to maintain because data is stored in a systematic manner and in a database. Decision making process would be greatly enhanced because of faster processing of information since data collection from information available on computer takes much less time than a manual system. The system designed and constructed in this project employing state of the art technology is at a reasonable low cost, enhancing that the functions the emergency assistance system already provides for the user can be done by using new advancements in technology. This approach makes it possible to overcome the shortcomings of the legacy systems developed before.

Even though the approach seems promising, it needs to be tested in the field using real time simulation and detection systems. One key advantage of this approach is that it only requires the user to download and run the application on their smartphones without any extra equipment or cost. This system can be used anywhere without the need for expensive systems.

Integrating Google Maps directly into the desktop application will make the navigation part even easier than it is in the current systems. Development of cloud computing can be used to do the processing on the cloud and instantly gives the feedback to the user. This will alleviate the problem of having to do a massive amount of processing in the phone. New features can be added based on IoT and Internet of Everything.

ACKNOWLEDGEMENT

It would not have been conceivable without the generous support and help of numerous Individuals who were there throughout the development. I would like to express my unique appreciation and thanks to industry people for helping me by giving information about the process of the work flow, and everybody else who bolstered me to achieve the goals of the development.

REFERENCES

ANAND, J. & FLORA, T. A. 2014. Emergency Traffic Management for Ambulance using Wireless Communication. IPASJ International Journal of Electronics & Communication (IIJEC), 2, 43-52.

BURKUL, M., NADMANE, A., SWAPNILUDAMALE & SHARADVAGARE 2015. EMERGENCY MEDICAL HELP SYSTEM USING OSM. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4.

HASAN, K. S., RAHMAN, M., HAQUE, A. L., RAHMAN, M. A., RAHMAN, T. & RASHEED, M. M. Cost effective GPS-GPRS based object tracking system. Proceedings of the international multiconference of engineers and computer scientists, 2009. 18-20.

JADHAV, SATAM & SALVI 2016. An Android-Based Emergency Alarm Message and Healthcare Management System. International Journal of Advanced Research in Computer and Communication Engineering, 5.

SARKAR, S. 2016. Ambulance Assistance for Emergency Services Using GPS Navigation.

ARGENTE, E., JULIAN, V. & BOTTI, V. MAS modeling based on organizations. International Workshop on Agent-Oriented Software Engineering, 2008. Springer, 16-30.

CENTENO, R., FAGUNDES, M., BILLHARDT, H., OSSOWSKI, S., CORCHADO, J. M., JULIAN, V. & FERNANDEZ, A. An organisation-based multiagent system for medical emergency assistance. International Work-Conference on Artificial Neural Networks, 2009. Springer, 561-568.

DAS, R. C. & ALAM, T. Location based emergency medical assistance system using OpenstreetMap. Informatics, Electronics & Vision (ICIEV), 2014 International Conference on, 2014. IEEE, 1-5.

GOUR & BAGHEL 2013. Gsm based heartbeat monitoring system

HAMEED, S. A., HASSAN, A., SHABNAM, S., MIHO, V. & KHALIFA, O. 2008. An efficient emergency, healthcare, and medical information system. International Journals of Biometric and Bioinformatics, 2, 1-9.

HASHMI, N., MYUNG, D., GAYNOR, M. & MOULTON, S. A sensor-based, web service-enabled, emergency medical response system. Proceedings of the 2005 workshop on End-to-end, sense-and-respond systems, applications and services, 2005. USENIX Association, 25-29.

JADHAV, R., PATEL, J., JAIN, D. & PHADHTARE, S. 2014. Emergency Management System Using Android Application. International Journal of Computer Science and Information Technologies, 5, 2803-2805.

JYOSTHNA, B. & GIRIPRASAD, R. 2013. Medical Emergency Assistance System using GSM. International Journal of Scientific Engineering and Research, 1.

MAHARAJA, R., REVATHI, S. & SUGANTHI, S. 2014. Gprs Based City Area Monitoring Using Embedded System. International Journal Of Scientific& Technology Research, 3.

MAURYA, K., SINGH, M. & JAIN, N. 2012. Real time vehicle tracking system using gsm and gps technology-an anti-theft tracking system. International Journal of Electronics and Computer Science Engineering. ISSN, 2277-1956.

RAO & KUMAR 2013. Providing Accident Detection in Vehicular Networks through OBD-II Devices and Android-based Smart Phones. International Journal of Science and Research, 2.

TAN, T.-H., GOCHOO, M., CHEN, Y.-F., HU, J.-J., CHIANG, J. Y., CHANG, C.-S., LEE, M.-H., HSU, Y.-N. & HSU, J.-C. 2017. Ubiquitous emergency medical service system based on wireless biosensors, traffic information, and wireless communication technologies: development and evaluation. Sensors, 17, 202.

UFOAROH, S., ORANUGO, C. & UCHECHUKWU, M. 2015. Heartbeat monitoring and alert system using GSM technology. International Journal of Engineering Research and General Science, 3, 26-34.

VORA, D., SHAH, R., SAWANT, S., NAIK, R. & D'MELLO, P. S. 2017. Emergency Management System Using

Android Application. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 6.

YUVARAJ, N., PRAKASH, V. & VENKATRAJ, D. 2011. Hi-Fi Traffic Clearance Technique for Life Saving Vehicles using Differential GPS System. World Academy of Science, Engineering and Technology.

Poster Sessions

ADAPTIVE SOLUTION FOR KEY CHALLENGES IN INTERNET OF MEDICAL THINGS

RMPHK Rathnayake[1] and MS Karunarathne
Department of Computing & Information Systems, Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, P.O. BOX 02, Belihuloya
[1]*rathnayakepiumi19@gmail.com*

Abstract- Internet of medical things refers to the worldwide network of interconnected medical devices based on a standard communication protocol. Moreover, it is about interconnected medical devices via the internet at any time, with anyone, at any place, to any service, from any network. With the rapid advancements of technology connected through the internet, the healthcare field also affected immensely. This study is an attempt to investigate the most useful technologies and key challenges regarding the Internet of Medical Things (IoMTs) nowadays. This paper proposes an adaptive model to address the identified challenges. The main contribution of this study is an entrusted framework for IoMTs which satisfies major challenges of security, privacy and the data integrity of the sensor data. Furthermore, the proposed cloud-based health management system will increase data availability, storage needs and processing power. In the cloud-mobile architecture, the security is entrusted with three methods: Advanced Data Encryption, Attribute-Based Encryption and Proven Data Possession. The proposed model shows how these 3 methods along with cloud technology address the identified challenges: Security, Privacy, Data Integrity, Processing Power and Storage Issues in medical applications. Local databases are the most common use of the data storage. In this model, it is giving the cloud-based solution along with the algorithm which helps to increase the security level of the sensor data. The used encryption and data provable methods are most recognized and strong algorithms in today world.

Keywords- Internet of Medical Things, Advanced Data Encryption, Attribute-Based Encryption, Proven Data Possession, cloud technology

I. INTRODUCTION

In the healthcare field, there are significant applications of the internet of things (IoT), such as medical equipment and medical medication control, medical information management, telemedicine, mobile medical care and personal healthcare management systems. These technologies are helping to make hospital intelligent treatment by collecting, handling, storing, transmitting and sharing digital data such as medical information, personal information, medication information, and equipment information within the hospitals.

In this paper, some of the major technologies of the internet of medical things are being evaluated.

A. Wi-Fi

Wi-Fi is a facility which allows devices to connect with another, though the internet especially without wired connections, in a particular area. Hospitals may have existing Wi-Fi infrastructure that allowing long communication range which is reducing initial costs. Wi-Fi can use in combination with other technologies. Wi-Fi may have higher power consumption too. (Jin, 1--2)

Wi-Fi networks are more vulnerable to attack by unauthorized users because they are more difficult to secure than a wired network. It also has installation difficulties as it is so commonly used. Consumers may find another Wi-Fi setup in the medical building interference with the wireless signal and it has law transmission speed

(Wells, 2009). Newly founded multi-Gbps communication is the evaluation of Wi-Fi technology which is to enable high-speed device synchronization (Cordeiro, 2010). Nowadays, there is much Wi-Fi enabled blood, telemetry systems, mobile x-ray machines, IV pumps and glucose meters in the hospitals.

B. Zigbee

Zigbee is IEEE 802.15.4 based technology used to create personal area networks. It is a low power, short-range, low data rate, low cost, low maintenance wireless ad-hoc network (Li, 2010). It has the potential for use in wireless medical monitoring. There are also drawbacks such as taking time to send set of data by up to four times compared to other low power technologies, lack of communication directly with existing cellular and internet infrastructure, low data rate, neediness of specialized skills to operate ZigBee compliant devices, security issues, high replacement cost, time consuming to adapt fast to the environment (Kanwar, 2012).

Zigbee technology is widely used in chronic disease monitoring, personal wellness monitoring, and personal fitness monitoring applications. Also, it can be used in devices that already exist including the pulse oximeter,

blood pressure monitor, pulse monitor, weight scale and glucose meter (Ahamed, 2009).

1) Zigbee System Architecture

operations also control by this layer. All the link setup, authentication, link configuration and other protocols carry by link manager protocol. Host controller interface allows command line access to the base-band layer and

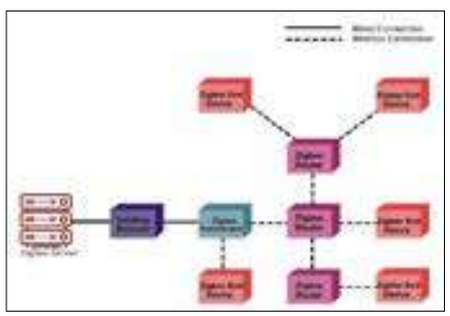


Figure 1. Architecture of Zigbee System

ZigBee system structure includes three different types of devices: ZigBee coordinator (ZC), Router(ZR), and End Device(ZED). Every ZigBee network should contain at least one ZigBee coordinator who acts as a root and links to the network. The coordinator is mainly responsible for handling and storing the information while performing the receiving and passing data operations. Zigbee routers act as middleware devices that enable data to pass to one to another device. End devices have limited functionality to communicate with the parent nodes (Alliance., 2006). The number of routers, coordinators and end devices depends on the type of networks such as the tree, star, mesh networks and so on (Horowitz, 2007).

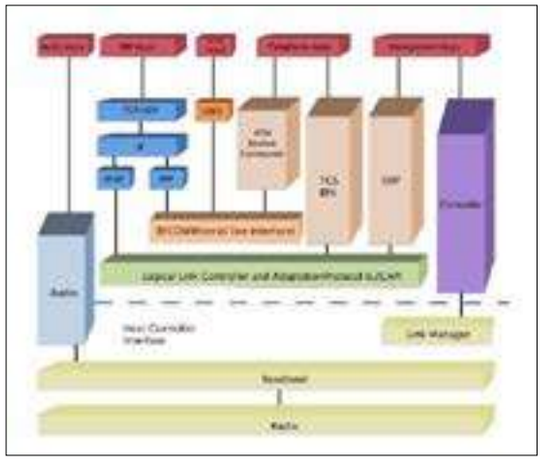


Figure 2. Architecture of Bluetooth

link manager for receiving status information. Protocol types are initially identified in logical link control and adaptation protocol (L2CAP). It provides connection-oriented and connectionless data services to host protocols above the L2CAP. Radio frequency communication protocol (RFCOMM) let upper layer protocols to communicate over a wired interface. So that applications do not need to know anything about Bluetooth. Service discovery protocol (SDP) uses to detect available services (Haartsen, 1998).

C. Bluetooth

Bluetooth is a wireless system that uses radio waves to communicate. It is a standard for the short-range wireless interconnection of devices such as mobile phones, computers, and electronic devices. The range of Bluetooth is 1 to 24Mbps (depends on version). The range of area is 30m. Bluetooth requires high power consumption. This is unsuitable for continuous medical monitoring applications (Bhagwat, 2001). Bluetooth is specially used in medical products such as Blood Glucose (BG) meters, Pulse oximeters, heart rate monitors and asthma inhalers.

1) Architecture of Bluetooth
When considering the architecture of the Bluetooth, it consists 2 layers: radio layer and base-band layer and 4 protocols: link manager protocol, logical link control and adaptation protocol, radio frequency communication protocol and service discovery protocol. These layers and protocols are connected via a host controller interface. All these functions run over the radio layer. Base-band layers work with the link manager to carry out the processes like creating link connections with the devices. Power saving

2) Bluetooth Law Energy (BLE)
Bluetooth Low Energy (BLE) is a feature which comes along with the Bluetooth version 4.0. The purpose of using BLE in healthcare systems is transmitting very small packets of data at a time while consuming less power. It is enabling Bluetooth smart devices to operate for months or even years. Because Bluetooth cannot use for continuous monitoring systems. BLE can be used for wireless devices to monitor and send medical information from biomedical sensors to smart-phones.

This technology has reduced the power and cost than Bluetooth. The range of the BLE is 15 to 30m. Bandwidth is 1Mbps (four times than ZigBee). But BLE is still an unproven technology. BLE implementation has 2 types such as single mode and dual-mode.

D. Radio Frequency Identification (RFID)

RFID is using together with the wireless sensor network in healthcare devices, especially for monitoring systems. These sensors are using to monitor physical or environmental conditions such as temperature, sound, vibration, pressure, humidity. Motion or pollutants at different locations.

Together RFID and wireless sensor network can use for remote monitoring of patients at home, collecting information such as blood pressure, heart rate, temperature, glucose level, analysing and making decisions. When trying to read RFID tags metals and liquids can cause problems if devices are in the vicinity of the distribution centers such as forklifts and walkies-talkies interruption may occur (Ajami, 2013). Use of the RFID in the healthcare industry alone will be faced with various issues.

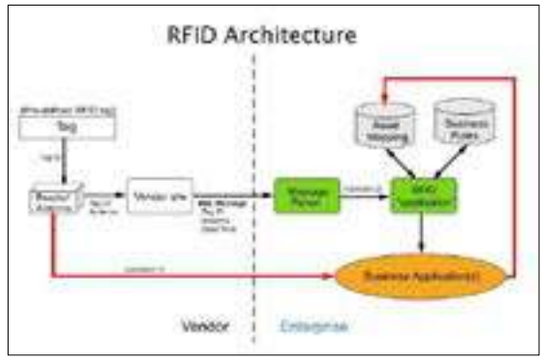


Figure 3. Architecture of RFID

Initially, item-tags are scanned by the Reader. Secondly, back-end transmitted data which are coming from an antenna. The signals are coming as RF waves. They have been recognized by the RFID-based system. The system acts as a middleware communication gateway among items, Reader, and the Databases. In the final step, it filters out and stores data in databases for checking the data faults and relevant operations.

E. Wireless Sensor Network (WSN)

A wireless sensor network (WSN) is a network which consists by a large number of sensor nodes where each node is equipped with a sensor to detect physical environmental things such as pressure, temperature, light, pH value. Today WSNs include sensor nodes, actuator nodes, gateways, and clients. The number of sensor nodes is deployed randomly inside of or near the monitoring area assuming the sensor nodes transfer the collected data to one another. During this transmission process to get to the gateway node, monitored data may be handled by multiple nodes. After multi-hopping, finally, it reaches the management (main) node through the internet or satellite (Gavrilovska, 2010). The one who handles the management node is publishing monitoring missions and collected monitored data.

Table 1. Technology comparison

Technology	State	Range (m)	Current rate	Bandwidth (Mbps)	Sharing			
					Data	Audio	Video	Voice
Wi-Fi	no	32	M	11	Y	Y	Y	VoI P
Zigbee	yes	100	-	0.2 5	N	N	N	N
Blueto oth	no	10-100	M	0.8	Y	Y	N	N
BLE	yes	15-30	-	1	N	N	N	N
RFID	no	1	L	1-11	-	-	-	-

Medium – M Large – L Yes -Y No - N

As other technologies, WSN also has security issues. Hackers can easily hack the network because WSN is working with nodes, they need to be charged at regular

intervals. So, the battery life of the nodes may very low. Like Wi-Fi and ZigBee, this technology also has low communication speed. And also, wireless sensor networks keep distracting from other wireless devices.

F. Cloud Computing

Cloud computing is widely used for delivering the applications over the internet as services. It is rooted in the internet search engine platform. It is an elastic resource which is scaled up or down effectively and efficiently. Also, a metered service so that consumers can pay only for what they use. There are three kinds of services in cloud computing such as IaaS (Infrastructure as a service), SaaS (Software as a service), PaaS (Platform as a service) (Armbrust, 2009). In the healthcare industry, cloud computing is used for population health management, care management and diagnostic, image handling services and laboratory services.

G. Global Positioning System

Global Positioning System(GPS) is a satellite-based navigation system which made up of at least 24 satellites. By using GPS tracking systems, that will make it easier for law enforcement to recover stolen property. When using GPS on a battery operated medical devices, there may be a battery failure and may need an external power supply which is not always possible. Also, in some cases, the GPS signals are not accurate due to some obstacles to the signals such as buildings, trees and sometimes by extreme atmospheric conditions such as geomagnetic storms.

III. IDENTIFIED CHALLENGES OF IOMT TECHNOLOGIES

Considering the above discussed technologies, five common problems were identified as follows,

i. Security issues
ii. Privacy challenges
iii. Limited processing power
iv. Limited data storage
v. Data integrity issues

A. Security Issues

As a growing number and variety of connected devices are introduced onto IoT (Internet of Things)-generation, the potential security threats are increased. IoT improves the quality of people's lives but also increases the potential attack surfaces for hackers and other cyber criminals. With the increase in the need for mobile systems, the current electronic market has also been filled with tabs, RFID devices, healthcare devices, laptops and many Wi-Fi enabled decides. Thus, the security of wireless networks such as Wi-Fi, Bluetooth, RFID, Zigbee has been increased. These wireless networks are prone to various attacks. The main reason for these security issues is that the wireless networks are open networks (Karygiannis, 2002). Most common security issues are,

i. Attack by hackers
ii. Intruders come into the network
iii. Data theft
iv. Virus attacks
v. Trojans and malware passed from one end-device to another
vi. Expose sensitive information to the open world

Security of wireless networks against such vicious attacks becomes the priority for the network industry. This is because not all networks are equally secure. The security depends on what security methods are used in the network, where and how this network is being used.

B. Privacy Challenges

There are privacy concerns in the wireless networks. The data transmitting in the wireless networks must be encrypted end-to-end device. If there is not proper encryption over the wireless network, sensitive data can be exposed to the outer world. The data that transmitting over the wireless networks should be stored securely, or else intruders and unauthorized persons can access the data, thus compromising the privacy of the users. Also, in actions like wireless hacking or some malicious attacks, if the sensitive information like patient records, patient contact details, payment information, history records, user identity, etc. are exposed or stolen, the privacy and confidentiality will be compromised.

C. Processing Power

Considering the transmission speed of the above technologies, Wi-Fi, Bluetooth, ZigBee, RFID, and WSN have a low rate of data transmission. ZigBee takes time to send a set of data by up to four times compared to other mentioned technologies. Most of these technologies access to the internet using access point (AP). However, the system has a difficulty of communication due to the traffic overload caused by communication through AP. (Jin, 1--2)

D. Data Storage Issue

Data storing is a very important fact in the healthcare sector. Because, the medical data are sensitive and confidential, hence the enhancements to privacy and security should be required. Most of the medical applications use local databases and management systems to store their data. It is easy for the third-party attackers to enter the systems. So that, Security and privacy enabled cloud-based systems are encouraged to use in the healthcare sector.

E. Data Integrity

Data integration in healthcare is about collecting, auditing and monitoring data. It goes beyond simple reporting of an organizational performance. It is not a simple task to integrate all the patient medical records universally. So that, there is a necessity for a method to integrate the data in an effective and efficient manner.

IV. CRITICAL TECHNOLOGY ISSUES FOR IOT BASED MEDICAL DEVICES IN HOSPITALS

A. Lacking medical device control

Healthcare device adoption simply indicates the growth of Machine-To-Machine (M2M) and Machine-To-People (M2P) automation. With such progress of medicinal workflow, IT functions have to focus beyond continuous connectivity with key applications and clinical employees. For each healthcare devices on the network, there is an application data flow between application and the system.

Variableness into medical device connectivity, locations, capabilities, as well as designs of activity is essential for optimizing medical care. Also, this is important for optimizing the infrastructure and for both of those short- and long-term planning for medical device automation.

B. Managing device diversity and interoperability

There are a number of healthcare devices are using in the medical field for different reasons. So that, healthcare stakeholders have to manage the different devices according to the different situations which create the need for technology experts in hospitals and medical centers. So before using medical devices they can give a training to the stakeholders who are going to use the device so that they can reduce the number of mistakes can happen. Not like normal mobile healthcare applications, there are huge devices like ECG monitoring devices, IoT based medical healthcare monitoring systems, ubiquitous medical healthcare systems.

C. Need of medical expertise

For some advanced medical devices, they need expertise in that area. In that case, the number of healthcare stakeholders can be reduced. So that it will optimize the flexibility of the device.

D. Hardware implementation issues

Hardware implementation part is the costliest in the healthcare domain. There are a number of sides in implementation they have to think like there are sensor-based devices, scanning devices which should have best quality materials, technology for hardware implementation, facilities to use those technologies, experiments, different conditions to use those devices in hospitals.

V. CONCEPTUAL MODEL

In this research, it has proposed a model for overcoming these issues in those technologies that are used in many healthcare applications and systems. In this proposed model, it uses Cloud Computing as a method to solve the storage capacity and processing power issues. And then to address the privacy issues, the proposed model is using a method called Attribute-based encryption (ABE). To overcome the security issues, the proposed model uses the Advanced Encryption Standard (AES) to encrypt the data before sending and storing in the cloud. And finally, to ensure the data integrity of the medical data that are collected from those technologies, the proposed model is using a method called Provable Data Possession (PDP).

Figure 4. Proposed conceptual model

VI. ABOUT THE ALGORITHM

The algorithm that implemented to overcome the data integrity, security and privacy issues in cloud computing has five main parts.

i. Advanced Encryption Standard (AES) Encryption
ii. Attribute-base Encryption (ABE) Encryption
iii. Advanced Encryption Standard (AES) Decryption
iv. Attribute-base Encryption (ABE) Decryption
v. Provable Data Possession (PDP)

The figure 3 illustrates the proposed security mechanism. In the first part of the algorithm, the data (text, image, sound, video and etc.) is encrypted using a symmetric cipher the Advanced Encryption Standard (AES). Then, the encryption of files is created using AES and generate AES-key. In this process, the AES encryption security level will take according to the specified security level in the algorithm which can be minimum (128 bits), medium (192 bits) or high (256 bits).

After encrypting the data and generating both the AES key and asymmetric cipher of the raw data using the AES encryption algorithm, this AES-key is used to the ABE Encryption. The AES-key is protected using ABE encryption. Here, we are using the CP-ABE (Ciphertext-Policy Attribute-Based Encryption. CP-ABE uses different pairing constructions to match the AES security levels of 128, 192 and 256-bits. In the CP-ABE encryption method, the data have associated with a set of attributes defined by the user. This creates an access structure policy for the encrypted data. The encrypted data is defined by a set of

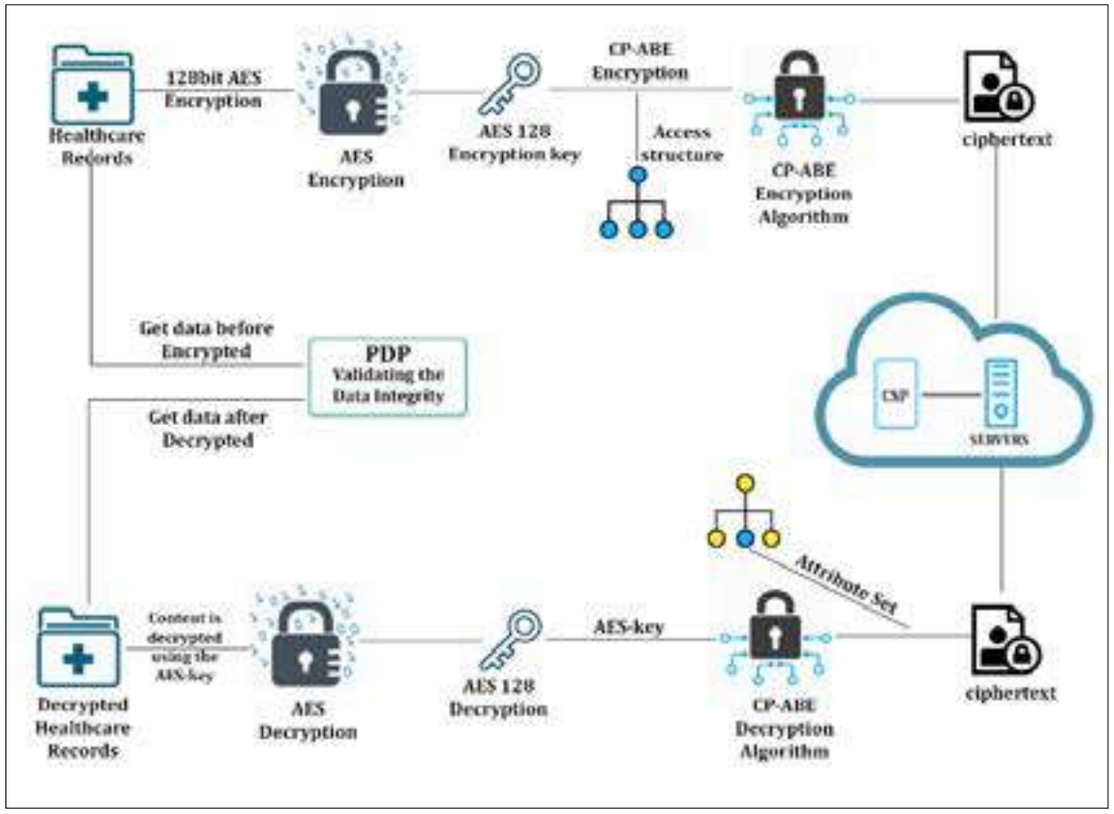


Figure 5. Proposed Security Mechanism

attributes, and access rule contained in the user's private key. If a set of attributes of data matches the structure of access to the user's private key, the data can be decrypted. After that, encrypted data will go to the cloud.

Once encryption is performed, the encrypted data (with AES) together with the encrypted AES-key (with CP-ABE) can be securely distributed over insecure networks or stored in an honest but curious untrusted third party (i.e. Cloud storage provider). Thus, the algorithm allows to achieving confidentiality and CP-ABE enable fine-grained control access mechanisms.

When decryption is performed, the AES-key is first decrypted using the ABE decryption key that matches the encryption policy. Then the content is decrypted using that AES-key. Only those authorized entities with a valid set of attributes could decrypt and recover the AES-key to launch the AES decryption process over the encrypted data. Under this solution approach, typical applications as securing digital medical records or storing and sharing of digital documents in the Cloud could be easily implemented.

Finally, the PDP method ensures the data integrity of the files that encrypted using the AES and ABE. This method enables a user to verify that the server possesses his data, without retrieving the entire data. Inside this method, it checks and validates whether the input raw data correctly store to the cloud and kept by comparing the encrypted stored data in the cloud a raw input data.

VII. CONCLUSION

The Internet of Medical Things (IoMT) is changing day by day. The world needs to prepare than with a value-based care. With the rapid growth of the IoMT the vulnerabilities of healthcare devices also increasing

day by day. This paper provides an overview of the IoT technologies, the key challenges in the healthcare sector and an adaptive solution to preventing the identified issues. There are five key challenges such as Data integrity, Security, Privacy, Storage issues and Processing power issues. The proposed model consists of 2 security methods such as Attribute-based Encryption (ABE) and Advanced Encryption Standard (AES) along with the data integrity method called Provable Data Possession (PDP). For the future work, it needs to enhance the effectiveness of the methods, as to deliver convenient usage to medical stakeholders. Further, we will look to use Artificial Intelligence-based encryption cryptography algorithms such as Neural cryptography to achieve more robust and efficient encryptions for data security and privacy of the healthcare organizations.

ACKNOWLEDGEMENT

This research is supported by Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka. And I am grateful to my supervisor Dr.(Mrs) M. S. Karunarathne for her immense support and guidance. I would like to thank my parents and friends who supported me to complete this research successfully.

REFERENCES

Ahamed, S., 2009. THE ROLE OF ZIGBEE TECHNOLOGY IN FUTURE DATA COMMUNICATION SYSTEM.. *Journal of Theoretical \& Applied Information Technology*, Volume 5.

Ajami, S. a. R. A., 2013. Radio Frequency Identification (RFID) technology and patient safety. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, p. 809.

Alliance., Z., 2006. ZigBee Security Specification Overview. *ZigBee Alliance. Wireless Control That Simply Works.*

Armbrust, M. a. F. A. a. G. R. a. J. A. D. a. K. R. H. a. K. A. a. L. G. a. P. D. A. a. R. A. a. S. I. a. o., 2009. *Above the clouds: A berkeley view of cloud computing*, s.l.: Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley.

Bhagwat, P., 2001. Bluetooth: technology for short-range wireless apps. *IEEE Internet Computing*, pp. 96--103.

Cordeiro, C. a. A. D. a. P. M., 2010. IEEE 802.11 ad: Introduction and performance evaluation of the first multi-Gbps WiFi technology. In: *Proceedings of the 2010 ACM international workshop on mmWave communications:* from circuits to networks. s.l.:s.n., pp. 3--8.

Gavrilovska, L. a. K. S. a. M. V. a. S. I. a. T. R., 2010. Application and multidisciplinary aspects of wireless sensor networks: concepts, integration, and case studies. In: s.l.:Springer Science \& Business Media.

Haartsen, J. a. N. M. a. I. J. a. J. O. J. a. A. W., 1998. Bluetooth: Vision, goals, and architecture. *ACM SIGMOBILE Mobile Computing and Communications Review*, pp. 38--45.

Horowitz, M. D. a. R. J. A. a. J. C. A., 2007. Medical tourism: globalization of the healthcare marketplace. *Medscape General Medicine*, p. 33.

Jin, C. a. C. J.-W. a. K. W.-S. a. Y. S., 1--2. Wi-Fi Direct data transmission for wireless medical devices. In: *Consumer Electronics (ISCE 2014), The 18th IEEE International Symposium on.* s.l.:s.n., p. 2004.

Kanwar, A. a. K. A., 2012. ZigBee: The new bluetooth technology. *ZigBee: The new bluetooth technology*, pp. 2319--7242.

Karygiannis, T. a. O. L., 2002. Wireless network security. *NIST special publication*, Volume 800, p. 48.

Li, J. a. Z. X. a. T. N. a. S. J., 2010. Study on ZigBee network architecture and routing algorithm. In: *Signal Processing Systems (ICSPS), 2010 2nd International Conference on.* s.l.:s.n., pp. 382--389.

Wells, J., 2009. Faster than fiber: The future of multi-G/s wireless. *IEEE microwave magazine*, Volume 10.

ANALYSIS AND DEVELOPMENT OF THE MESS MANAGEMENT SYSTEM FOR THE KDU CADET MESS

ODS Fonseka, RL Jayakody[1], PHP Arachchi , and SCM de S Sirisuriya
Faculty of Computing, General Sir John Kotelawala Defence University Rathmalana, Sri Lanka
[1]ravinajay99@gmail.com

Abstract- Today's Sri Lanka Military Mess Management and cost calculations are done manually and there is no proper system with electronic payment handling. Hence all the difficulties in manual system such as time consumption and human errors are inheriting to mess systems. Due to these problems, a requirement arises to create a system that will make the entire Mess Management an automated system. The present work automated the mess system which allow manipulate all the details about the payments and military personal and easy generation of reports using C# and SQL server. Further system also provides data on bar sales to be analysed with use of dashboards. Finally the present work allows mess management staff to engage in their work with maximum security and risk free from errors or loss of data. This new automated system will be interconnected with all end-users of the mess and all duties will be able to be proceeded easily with this Mess Management System.

Keywords- Stock, backup, restore, add, delete, Payments, Reports

I. INTRODUCTION

The aim of this project is to develop an accurate and efficient and computerized cadet mess management system for the KDU which can overcome the problems and loose ends of the existing way of managing and monitoring the information such as data entering issues, reduce delays in searching records, and handle all the related activities of the cadets to provide maximum service to users.

Having a variety of food within the menus for foreign and local cadets, the system will be able to felicitate the required menus payment calculations requested by the foreign and local cadets. The mess in KDU is compromised with cadet officers in three types of locations such as cadets in Rathmalana, cadets from Werahara and cadets from Sooriyawewa. Each section consists of either local or foreign cadets respectively.

The main feature of this project is to generate a hardcopy receipt of the total expenditure of the cadets at the end of each month. A special area will be allocated to get the necessary information about monthly expenditures. The payments can be paid by the end-users of the mess with the monthly invoices issued at the end of each month through the automated new mess management system.

More other features which are included in this computerized mess management system (MMS) is that a report will be produced at the end of each month consisting with the total monthly expenditures and other necessary information. The reports will be separately generated by the mess management system to the relevant sections consisted and letters will be generated in order to make the system completed with more efficiency directly through the software application than handwriting which is a waste of time and increases the expenses.

The main problem aroused was the inefficiency of the currently prevailing manual mess management system. It is time consuming and even the slightest mistake will effect errors in any money transaction details or monthly bill details. The misplace of data and information will always occur in such manual systems and data entry and

retrieving will take a considerable time due to the absence of an automated system. No authorized method for mess managers to keep in touch with the standard of security within the manual system is also a major problem. Less protection methods since it is hard to have more back-up copies which will again result in using more storage space and therefore increase expenses.

Through such a mess management system the aims such as organized in time management, security, stock level management, data storage, information gathering, report generating, searching for details, errors in calculation, backing-up and trustworthiness can be achieved.

Using a user friendly system which consists of user interactive interfaces with the use of suitable technology is another objective through this mess management system. Delivering the system with the maximum functions in order to make the system functional within the management of the mess is the main objective and also to enhance backing up the entire system to get rid of the loss of data.

II. LITERATURE REVIEW

When considering the work done in the previous mess management system the whole system is driven manually with manual calculations and data storage in previously named and arranged files within the specific section. Considering the KDU mess, the mess duties which should be computerized can be identified into two main sections such as the payment handling with cadet details and bar section.

The sections are managed and data is organized by the specific officer assigned in order as to do his or her duty in that section. Considering the cadet and officer mess detail entry, deletion and updating, the current system enters the details into spreadsheets and those are taken with printouts to store in a separate file so needs in preparing the mess bills separately to each user individually with the expenses within the month. The monthly generated bills are pasted in a large book which acts like a database but it is a time consuming and complex process than the easy to use featured user interface and data entry interfaces in the new system. There is no feature to search and check required cadet details through featured buttons or options and one by one should be analysed in order to get a cadets details. Each cadet consists of a serial or registration number which will be the main attribute in order to recognize that this is the particular person which an expense should be added to.

In the payment section the payment details are made as in for issuing the mess bill for the services and food provided by the mess to the end-users of the mess. The payments include all the necessary payment criteria's for a month and additional payments which are only included if a particular event is held in a specific month. The whole receipt is a printed bill, but all other calculations are done in a separate place not attached to the system directly.

In the bar section of the mess there are two main criteria's followed in the current system. They are to note down the stocks of sales and the purchases to balance the remaining amount of the stock. The current bar section in the mess is not interconnected with the payments directly and due to that two people are working separately. It is a disadvantage to maintain records if one administrator is unavailable in one section. Therefore, through the mess management system, both sections will be interconnected with one major system where information can be shared directly with the two administrators using the system.

A system was developed using the Model-View-Controller architecture which divides the programming structure into three tiers known as model, view and controller. Model governs the access to and updates of the system data. In real world, the model could be represented as the business domains. Since this system is developed using Java language, therefore, NetBeans IDE is chosen as it is designed for Java application development. NetBeans IDE is free and open-source, so, there is a lot of information about it available on the Internet and it is convenient to find support for it. The system is backed up using internet cloud applications and security measures are taken because the system is user's servers and web technologies in inter connecting the system. (Muniraja, Rajanikanth, In-Time Billing Process for Canteen Management System)

The functions and features of a previous mess management system are as follows. The mess manager will first login by entering the correct username and password. The user will need to enter the correct old password in order to set a new password. After successfully logging in the options page will be displayed with following options such as Cadet-info, Stores, Market rate, Diet-chart, help. (Ankita Chawla, Priyanka Joshi, Sanjana Panjwani, Surabhi Sontakke, Mess Management System)

In the navy department mess management system, it runs fully manually with the information and details recorded in files and books. It consists of payment handling books and Common Cash Card (CAC) automated system. In the Cash Meal Payment Book records on meals sold for cash are recorded. When meals are sold for cash from a general mess they give a receipt through transfer Control process. The supply officer will assign a control officer for the handling and security of the Forms. Assignment will be as a collateral duty and shall be listed in the command notice of collateral duties. The Transfer Control and Receipt will be used to complete books. Individuals authorized to receive cash meal payment books will sign the transfer control and receipt at the time of receipt. The coupon will be retained by the control officer transferring the book as a receipt. A transfer control and receipt will be used to return the completed books. The transfer control and receipt coupon will also be used when the Cash Meal Payment Book is turned over to station audit boards. (Michael S . Hansen CDR ,SC , USA- NAVSUP P-486)

III. METHODOLOGY

A. Description for proposed system

Considering the above mess management systems and other related systems which consist of stock and payroll management systems we propose a more efficient and functional mess management system called MMS for the KDU mess for the administrators to use it easily and to secure details of end-users.

The proposed system will be more user friendly and interfaces of it will be more specific enough to give all directions to the user to proceed with all functions easily. The proposed system to KDU will be run on the program application and all user requirements which will be needed for the payroll and stock management systems. The whole model will be more efficient for data backups, insertion, updating, deletion of information and generating user required reports. Visually providing all detailed paths, having support from a guide in order to use the system wisely and providing the opportunity for the administrator to have easy access from forms or girds are other features provided by the proposed mess management system.

The overall process of the Mess Management System has been implemented as a computerized solution to run within the mess office. It works on a program application based software using C# and SQL server. The activities in the system will serve upon the inputs and request of the user. The system first will be open with a splash screen. After entering the splash screen the login screen will be appeared. The login screen is the first user interface which the end-user will have to interact with the system to proceed relative functions and duties.

In the development of the Mess Management System to the KDU mess, it is planned to use the windows based technology. The use of the windows based technology is due to the reason because, after the development of this system, this system will be used only by a single end-user in the mess. The programming language that is discussed to be used for development is because the Mess Management System is highly dependent on the accuracy and efficiency.

B. Functional requirements

In order to build up a more efficient and time saving mess management system to manage all information on the users of the mess the best software solution is to have a computerized mess management application program which has all the functions in order to proceed with day to day transactions done within the mess and generating the necessary monthly bills or reports for the mess.

- Application should be able to provide secure authentication.
- Application should be able to store data and to maintain the database.
- Application should be able to get details of the users of the mess.
- Application should be able to get details on the payment amounts.
- Application should be able to get details on the stock amounts.
- Application should be able to calculate the total expenditure and balance of each individual user.

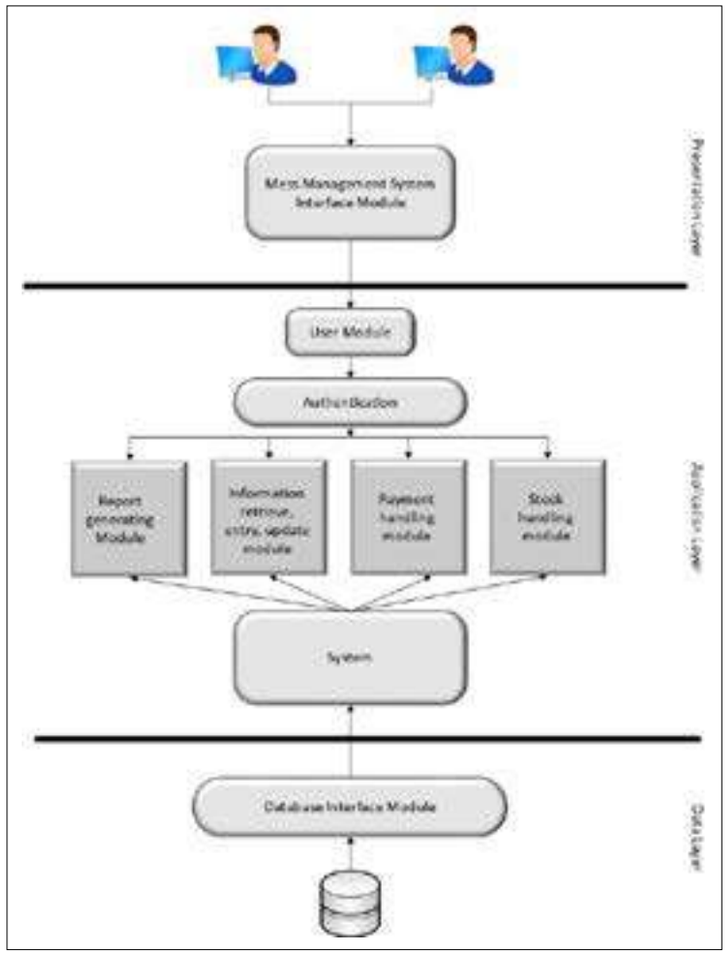


Figure 1. Top level diagram of proposed system

- Application should be able to generate the monthly mess bill for each user.

C. Non-Functional requirements

Non-functional requirements statement features of the system other than the detailed functions it performs. These features include system performance, speed, security, reliability and time management system characteristics. The non-functional requirements also talk aspects of the system development process and in use personnel. It includes the following:

- The system will not allow unauthorized accesses.
- The system will not allow users to access pages which permission not granted to access.
- The system will enable efficient management with search and retrieval of data records.
- The system should be able to reduce the time that take to process a record than in the current system.
- The system will allow the data stored to be available for updating or deletion.
- The system will be able to protect from data losses through backing up data.

- User friendly interface.
- Details on each button and functionalities should be provided within the interface.
- Easy to handle with and consisting a help guide.
- Changing skins.
- Proceed with generating monthly mess bills for all users in a less time.
- Perform calculations quickly on payments and stock.
- Data needs to be processed in a minimum required time limit.

IV. HOW THE SYSTEM WORKS

The activities in the system will serve up on the inputs and request of the user. The system first will be open with a splash screen. After entering the splash screen the login screen will be appeared. The login screen is the first user interface which the end-user will have to interact with the system to proceed relative functions and duties.

A. Login section

The cadet mess login screen will be processed for correct authentication to enter the system and handle it. The end-user will be allocated a specific username and password to be entered and if any error occurs in invalid usernames or passwords the system will automatically deliver an error message. The error message will be generated for incorrect usernames or passwords separately and if both are not entered a message box will be shown that user should enter it properly.

The user authentication is given in order to secure the details from out siders who can make changes in the mess management system without any authorization. In the same screen the end-user will get the option to retrieve the password he or she forgets it by correctly providing the emails details to get back the login details to re-enter the system.

B. Main User Interface (ribbon form view)

After correct Verification of user name and passwords the end-user can access the cadet main menu screen. The main screen will be consisting of a ribbon type screen as shown in Figure 1 which will be made from the DevExpress User Interface designer. The ribbon will be grouped into 4 main sections as the end-user requested to include in the system on the modules such as cadet, bar data and application configuration. In each section it consists of sub sections which perform the usability tasks within the system.

C. Cadet Details Section

Cadet details option the end-user will be given an interface to enter cadet details according to the requested attributes of the mess management staff. The cadet details will be added under sections such as whether it is KDU Rathmalana or KDU Sooriyawawa or KDU Werahara, degree programs and respective troops. An additional grid view will be also provided to end-user to view all the cadet details which were previously added by the end-user. The cadet's details can be searched with the search option given and relative information is given in the grid.

A row of the grid the end-user can access the selected information from the text boxes provided to enter details. This is to access information accurately without any end-user mistakes in reading data through the grid. The same form will give another option to view the grid in a separate form in order to enter, delete or update details within the system as the request of the end-users of the mess staff. Options to add, delete, update will be provided with buttons in the ribbon form.

In the grid view details of the cadet can be easily found from the find option above and details can be grouped as the user wants by dragging the columns under which preference needed to be viewed. The chosen preference details will be only displayed in that information section.

D. Cadet Payments Section

The payment section screen the details of costs, which should include in the mess bill at the end of the month are included. The form is run on the cadet's service numbers, intake wise expenses and main costs will be included it. If there are any changes to be done the end-user can choose the specific cadet and make changes in costs as in for the services taken

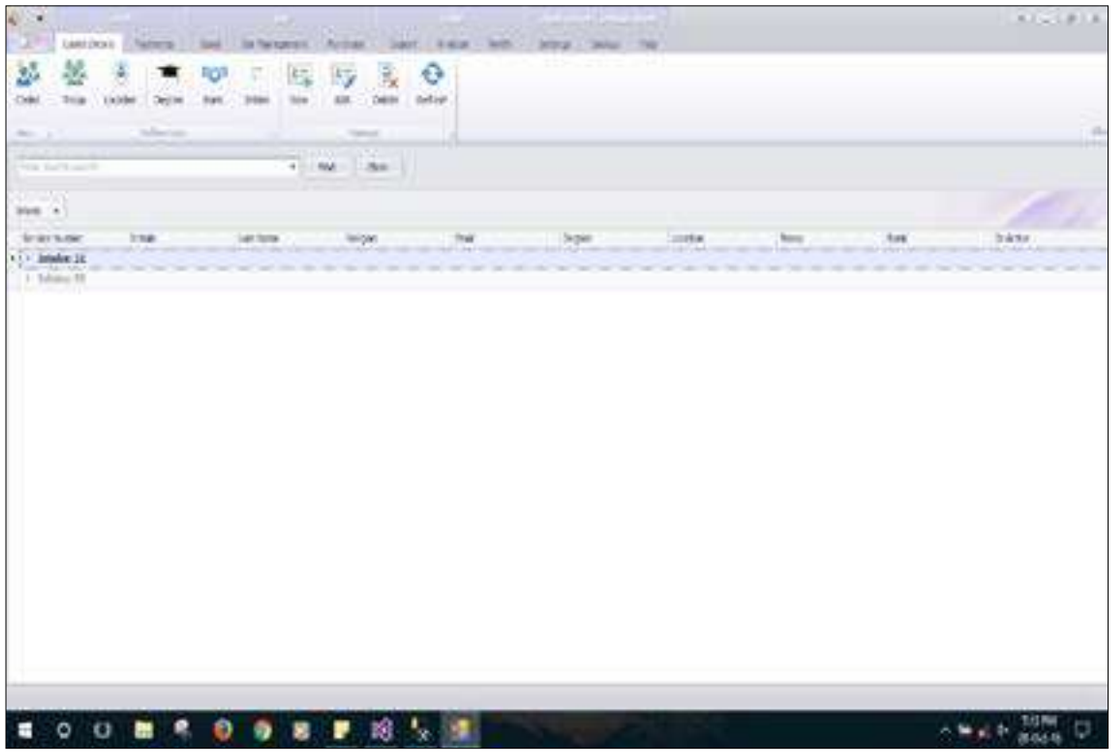


Figure 2. Interface of the main ribbon form

by that cadet for the month. Options such as add new, update, delete and view details will be included for the end-user on their request. The payment form will also give an option to view all payment details in a form of a grid for more user convenience.

In the additional payments form the additional details costs will be included. Additional costs are the costs which will be not always included in the monthly mess bill and if only that cost or service is made it will be included into the mess bill. The additional costs will be added to each cadet after searching the cadets' service number and for sudden events more empty text boxes will be included in order to enter the event name and costs amount. This form too will compromise with add and update in the additional payments detail form.

F. Bar Management

In the Bar management section it will be mainly about the adding of products and product categories for the system. products are added under the product category and each product can be edited, deleted or updated as the user requests in the mess management system. The products will be displayed in a grid for the user to display and view in any need.

G. Mess Stock Section

In the stock section the details of the stocks in the mess of all products are managed. The stock option will be under main functions. One for adding a purchased item to the stock and the other is to update any detail in the stock items. The purchased items for the stock are detailed with its Quantity, prices and expiry dates etc.

If a purchase is done this is added to the stock under a batch number system. in this non-functional requirement if the cost price, sales price and expiry date is different in the same products previously added details a new batch number is issued for the new batch to be entered to the stock details under a new batch number. If cost price, sales price and expiry date is same the purchased quantity will be added to the same product under its old batch number.

In the stock option in the stock section the stock details such as the cost price, sales price and expiry date can be edited on a certain product under its batch number. All details are provided to be viewed in a grid view of the user interface in the stock section.

H. Sales Section

In the sales section the details of the cadet's sales are included. The cadet should be displayed first and the product which he or she will be purchasing will be displayed with the respective batch number and the quantity. After entering all the sold items for the cadet the system will send its details to the reporting section to include it in the mess bill as required.

I. Data Section

This will be on the reporting and exporting of reports. The reports can be generated as needed by clicking the options given in the export button in the ribbon form.

J Application Configuration Section

There will be option for the users to change the skins of the Mess Management Systems user interfaces as required for easy sight preferences. The help option will give the access to the help menu where the end-user can get details and help on the whole mess management system

Even new end-users will be able to handle the mess management system with the reference of the help menu provided. The about the system option will provide a screen which gives the details of the developers and product license of the mess management system. In the Backup option it will be beneficial to backup all details of the system to a cloud. This feature will be provided by the engine will be important to the end-user to save all details of the database of the mess if any error or information is lost within the system. Therefore, through the backup option, all information is secure at any moment.

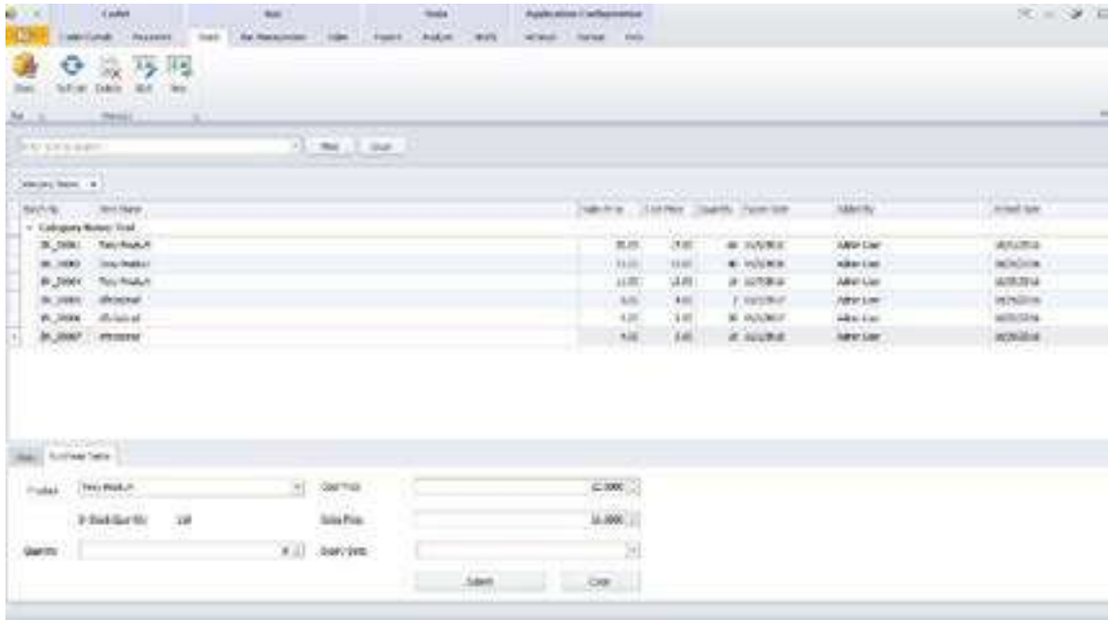


Figure 2. Interface of stock section

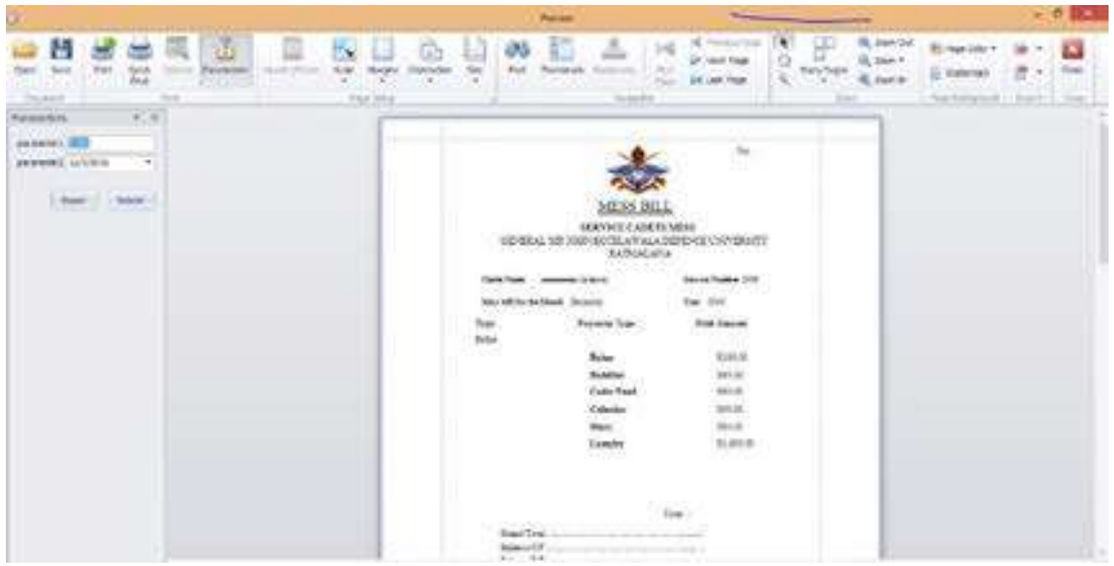


Figure 3. Mess bill preview

The update process of the system will be done by the administrators in charge of the system assigned by mess. They will keep in touch of the frequent changes of the system as required. The system is 24 hours accessible and therefore the system should be frequently updated. Within the process of this mess management system always keep track of the accounts within the system and other bill clearances.

V. DISCUSSION

The aim of this project was to develop an automated mess management system as a solution for manual mess management system of KDU cadet mess. The development team implemented this system in order to determine its ability to satisfy the entire functional and non-functional requirement with special qualities such as flexibility, reliability efficiency and etc., to overcome the drawbacks identified in the system.

The study found out that it is feasible to use the language C# in .NET frame work to develop the mess management system since it brings direct advantage of platform independency, automated monthly mess bill calculation, automated stock control system and report generation which were identified as main necessities to enhance the productivity of the cadet mess environment. Calculating monthly mess bills by manually using MS Excel, unavailability of separate function for taking backups of daily transactions, not compatible with new operating systems are some issues which are clearly answered by new computerized mess management system.

VI. CONCLUSION AND FURTHER WORKS

The computer based mess management system is able to gain various achievements which will satisfy user expectations. The user is provided the option of monitoring the records entered earlier and can get details of the users of the system. Data storage and retrieval will become faster and easier to maintain because data is stored in a systematic manner and in a single database. Also the system is able to backup all data and make data updates easily.

In result the system is able to provide secure authentication with using md5 authentication. The representation of interfaces with a more creative and user friendly way also will be beneficial to the mess management system. Having the required features in order to the mess management staff to do their daily data entries, updates and retrievals are also something which can be gained by the new Mess Management System.

As further work we are hoping to implement this system to the KDU officer mess and combine the two systems for easy management within the University for payment, stock and report generation purposes. Introducing new features according to the user requirement and their expectations. So as a future enhancement, hope to introduce the system in web based to coordinate the processes through the web. By implementing and facilitated the site with the web service it is possible to increase the usability of the system.

ACKNOWLEDGMENT

We want to thank all the lecturers of KDU IT Department and KDU mess who gave us very helpful feedback to complete this work.

REFFERENCES

Muniraja, Rajanikanth, In-Time Billing Process for Canteen Management System,ISSN 2247 3983.

Ankita Chawla, Priyanka Joshi, Sanjana Panjwani, Surabhi Sontakke, Mess Management System,International Journal of Research In Science & Engineering e-ISSN 2394-8299 Volume 1 Issue 2.

Navy department naval supply systems command, Michael S. Hansen CDR, SC, USA- NAVSUP P-486.

Eng wei seng, Wong yin cheng ,Cateen management and catering system.

EMERGENCY ALERT SYSTEM FOR REPORTING CRIME ISSUES TO NEAREST POLICE STATION

LC Wijenayaka[1] and N Wedasinghe
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]lahirucwijeneyaka@gmail.com

Abstract- In the recent past there had been a tendency of crime rates to increase and the main reason for this is the lack of possible means to keep the Police informed immediately. Thus there is a possibility of the criminal escaping the crime scene when the Police arrive. A mobile application with an Emergency Alert System to report crimes to the nearest Police Station has been proposed to overcome this issue. There are different types of Crime Alert Systems used all around the world. However in Sri Lanka there is no properly developed computer-based Emergency Alert System to report crimes to Police. The utilization of the introduced application will allow the witness or the victim to inform the nearest Police Station and Mobile Police. The application allows the user to mention the location and share the type or nature of the crime which will allow the Police Officer to be aware of the nature of the crime before he/she arrives at the place of crime. Using the tracking facility available in the application the Police Officer will also be able to identify the exact location of crime. Therefore, this application will assist in the process of decreasing crimes.

Keywords- GIS, Location Tacking, Policy, Crime

I. INTRODUCTION

Generally, the definition of crime is an act punishable by law, usually considered to be a bad act. Crime refers to many types of abuse prohibited by law. Crime includes such things as murder, theft of a car, arrest, possession or dealing with illegal drugs, being naked in public, drunkenness, and theft of the bank. Crime is an act that has been timeless and has been done practically since the beginning of time.

Crimes against people include assault, kidnapping, murder, and sexual assault. Such crimes usually bring serious penalties. Offenses against property include fire incidents, automobile theft, theft, embezzlement, forgery, fraud, negligence, and vandalism. In most cases, in these crimes, there is a slight penalty against the person against the specified crimes. Crime is the most difficult thing to commit. According to this case, the law treats the crime against theft or crime against the person. By using this mobile application anyone can inform the locations of that crime happens. Anyone can use this mobile application from anywhere by typing their name and their mobile number. And then he/she can track the locations and send it to the nearest police station. Then user will get a confirm message whether the tracked location and other information's are send to the police station.

Nowadays there is a possibility to increase crimes in Sri Lanka; there are no strong strategies to control those crimes. Thus because of limitations of informing crimes to police and because of people has no proper way to inform to police about crimes increase this conflict most manner.

Using this application, it happens a crime people can inform police immediately through informing current location to police. The message which generate this application basically send to nearest policeman mobile and nearest police station. In case of information confirm police, officers can arrive crime location in correct, manners in correct time. Basically, this application will help up to certain point reduce the crimes in Sri Lanka.

The initial aim of this project is support police to identify the exact places and aware the people about crimes. Generate a report regarding crime arias. Through this application possibility to reduce happening of crimes in Sri Lanka. We have identified the following objectives.

- To critically study the existing computer-based solutions for Crimes in Sri Lanka.
- To conduct a detailed study on current mobile applications, database and web technologies.
- To design the system and develop the prototype properly.
- Testing and evaluation of the new Emergency alert system for reporting crime issues to nearest police.
- To get the results of the evaluation criteria.

II. LITERATURE REVIEW

Through the review, it will be thought about regarding the procedure and available practices in crime issues and alert systems.

Jonathan, Klickjonathan.Klick (2003) has presented, using terror alert levels to estimate the effect of police on crime. Changes in the terror alert level set by the Department of Homeland Security provide a shock to police presence in Washington, D.C. Utilizing day by day wrongdoing information amid the period the dread ready framework has been set up, we demonstrate that the level of crimes diminishes essentially, both factually and monetarily, amid high-ready periods. The diminishing in the level of crimes is particularly huge in the National Mall. This gives solid confirmation of the causal impact of police on the level of crimes and proposes an exploration system that can be utilized as a part of different urban areas.

LI YUAN, (2007) has developed a system for this Problem by developing wireless messaging used to fight crime. At the point when police in The Hague in the Netherlands got a report that a vessel had been stolen, they conveyed an instant message about the case to inhabitants who had joined to get neighbourhood crimes cautions on their phones. After a hour, a lady bicycling along a channel who got the message told police through a telephone call that she saw a boat that met the portrayal. The boat was found, and the thief arrested.

Hartman, Ginger. (2012) has developed a system for this problem by developing Wisconsin Crime Alert Network. The Wisconsin Crime Alert Network from the Wisconsin Department of Justice permits neighbourhood, state, government, and ancestral law implementation offices to convey crime ready notices to organizations and people in general focusing on beneficiaries in light of sort of business and area. Law Enforcement Officers from all through the State of Wisconsin will convey cautions about crime and criminal suspects. The officer will choose which of roughly 50 gatherings ought to get the alarm. Gatherings will incorporate drug stores, comfort stores, and banks, and additionally associations and offices, for example, neighbourhood watch, hospitals and schools/ colleges, and private natives. Officers can likewise transfer photographs to send with the alarm. Next the officer will choose a geographic territory to get the alarm. Once the officer sends the caution it goes out immediately to beneficiaries.

Maguire, Ken. (2006) has developed a system, Crime alerts at your fingertips in Boston. Boston is launching a crime alert system that will send text messages, and faxes to inhabitants when crime happen in their neighbourhoods, police and city authorities said yesterday. The framework, keep running by the Boston police and the Internet organization CitizenObserver.com, is meant to disseminate crucial information about crimes, including times, areas, portrayals of suspects, and photos, under the control of those most influenced and those in the best position to enable police to discover suspects. Authorities said they would like to draw in inhabitants, particularly network and crime watch gatherings, as police battle a noteworthy upsurge in crime in some Boston neighbourhoods. By giving at times prompt data and routes for inhabitants to message tips back to agents, police would like to pick up a great apparatus in recognizing and getting offenders. The electronic tips that inhabitants can send back to police would be mysterious, conceivably liberating some from fears of reprisal for helping specialists. The program, which is likewise being utilized as a part of Cincinnati and in a few different urban areas the nation over, enables police to take the data they have gathered from a break-in or bank theft and post it on a site, where it is then consequently transmitted to anyone who login to get the messages. The alerts are also available online.

Witte, Ann Dryden.(1996)Research suggests that some social and criminal justice policies can affect the Crime rate. This is about major criminal justice and social policy issues related to urban crime, such as drugs, domestic

violence, property values, and the underground economy. Family disturbance, drugs, constrained financial openings, and vacant and unsupervised youth are altogether observed to be related with urban crime. The article presumes that real reductions in crimes are probably going to come about just from expanded monetary and social open doors for families and youth, especially for young males. Escalated programs coordinated at families and in danger youth will probably bring lower crime than are programs coordinated at individuals as of now vigorously associated with unlawful exercises. It costs less to keep youngsters in education and training programs than to detain them, and will probably create gainful and balanced grown-ups.

III. METHODOLOGY

A. Data Gathering

Qualitative and quantitative data required for designing the requirement specification for the new system were gathered by conducting a survey.

B. Approach

In the Emergency alert system for reporting crime issues to nearest police station, process first user should login to the Mobile application by providing user's name and phone number to the system. Then the mobile application will track the coordinates (longitude and latitude) of the location where user is in (where the crime happens). Then those information; User name, User mobile Number, Date and time, small message and coordinates of the tracked and will send to the web server that is in nearest Police station. After taking the information from the mobile application those details will be displayed on those web servers. The location of the crime will be display on a map according to the coordinates that sent from the mobile application. And the web application will send a notification for the user whether the information send successfully or not and give a compliment for the sender for his service Each information that users send will stores in SQL server database successfully and for the work with the application it required continuous internet connection.

C. Technology Adopted

It is necessary to use new technological methodology for the system. It is very important to use acceptable tools so as to develop productive system. Use on any inappropriate tools can solely ends up in develop a system with unnecessary errors and faults and use of those badly chosen technologies additionally can ends up in crashed when the new system implementation. Badly chosen technologies which can be extremely advanced and complicated will enable manufacturing a system with a top quality, however these technologies may result in develop a system that spend lots of time and resources so as to perform a task that is anticipated by the system. It is very important to use applicable programming language and the other necessary tools in order to develop a productive system. Therefore, these technologies and tools can help to develop the system among a minimum development time the most objective of developing this type of an application is to produce the users more efficient work system instead of doing manual approach. Because of that we should use the most applicable tools available in the market to develop the system. Technological considerations - followed during the development of the system Efficiency and Performance Re-usability and flexibility object-oriented development support so according to the Emergency alert system for reporting crime issues to nearest police station. Java and android studio used to develop the mobile application. According to that requirement system has developed by using C# and using SQL database to run on windows operating system.

• Web Application

The programming language that is going to apply as the developing language for the system development turned into significantly trusted accuracy, performance. When considering all these technologies which can be associated with the Emergency alert system for reporting crime issues to nearest police station can be applied a web-based technology. The .NET Framework comprises of the regular dialect runtime and the .NET Framework class library. The regular dialect runtime is the establishment of the .NET Framework. You can think about the runtime as an operator that oversees code at execution time, giving providing core services such as memory management, thread management, while additionally authorizing strict write wellbeing and different types of code exactness that advance security and power. The class library is a thorough, question situated gathering of reusable kinds that you can use to create applications extending from customary charge line or graphical UI (GUI) applications to applications in view of the most recent advancements gave by ASP.NET, for example, Web Forms and XML Web services'# programs run on the .NET Framework, it runs on a virtual execution system called the common language runtime (CLR) and a combined set of class libraries. The CLR is the implementation by Microsoft of the common language infrastructure (CLI) and it helps to create execution and development environments in which languages and libraries work together without any flaw. C# source code is compiled into an intermediate language (IL) that conforms to the CLI specification. The IL code and resources such as bitmaps and strings are stored on disk in an executable file called an assembly. It contains a manifest that provides information about the assembly such as types, version, culture, and security requirements.

When the C# program is executed the assembly is loaded into the CLR. Based on the information in the manifest CLR might take various. When the security requirements are met, the CLR performs just in time (JIT) compilation and convert the IL code to native machine code. CLR also provides other services such as exception handling, and resource management. It illustrates the compile time and run-time relationships of C# source code files and the .NET Framework. It demonstrates the relationship of the common language runtime and the class library to our applications and to the general framework. The representation likewise indicates how overseen code works inside a larger architecture.

• Mobile Application

Android gives you opportunity to put into impact your own gadget details and drivers. The equipment reflection layer (Hal) presents a notable approach for creating programming snares among the android stage stack and your equipment. The android working machine is in like manner open source, so you can make a commitment your own interfaces and upgrades. Android is an open source, Linux-based programming stack made for a wide exhibit of gadgets and shape factors. Android moreover comprises of an immovable of center run time libraries that offer the majority of the ability of the java programming dialect, which incorporates some java 8 dialect works that java API system employments. Applications, which broaden the usefulness of gadgets, are composed utilizing the Android programming advancement unit (SDK) and, frequently, the Java programming language. Java might be joined with C/C++,together with a decision of non-default runtimes that permit better C++ support. The Go programming dialect is likewise bolstered, in spite of the fact that with a restricted arrangement of use programming interfaces (API)

• Database Selection

Consistent with the above eventualities most of the structures are used square database to keep facts. It seems it is simple to control and perform. So, the database put in force on the server has to able to supply efficiencies operations. Consequently, the proposed system decided on the Microsoft SQL server as server. SQL server is the inspiration of Microsoft's data base platform, delivering challenge critical performance with in-reminiscence technology and quicker insights on any information, whether or not on-premises or in the cloud, and also Microsoft SQL Server is an application used to create computer databases for the Microsoft Windows family of server operating systems. Microsoft SQL Server provides an environment used to produce databases that can be accessed from workstations, the Internet, or other media too. Database management or DBMS, store user's data and enables them to transform the

information into statistics. Those systems allow users to create, replace and extract facts from their database.

A database is an established collection of information. Facts refer to the characteristics of human beings, things and activities. Square server stores every statistic item in its very own fields. In square server, the fields related to a particular character, thing or occasion is bundled collectively to shape a single complete unit of records, known as a document. Each record is made up of some of fields. No two fields in a record will have the equal area name.

Throughout an SQL server database design project, the evaluation of your project wishes identifies all of the fields or attributes of interest. If your commercial enterprise desires trade through the years, you outline any extra fields or alternate the classification of present fields.

D. Overall Architecture.

Overall architecture outlines general structure of whole software improvement. Overall system will be part into three layers as presentation layer, application layer and data layer. In light of this structure, the parts will be isolated into particular modules to deal with the operations of all parts. Overall architecture of the proposed system. General framework engineering of the proposed framework is given beneath. (Figure 1.)

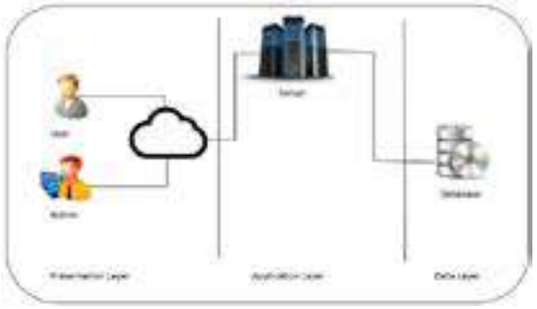


Figure 1. Overall Architecture

E. Functional Requirements.

- Mobile application should be able to track the location of the Place that Crime happen.
- Mobile application should be able to send the particular details (user details, user mobile no, date, details about crime)
- Web application should be able to get the information that sent from the mobile app.
- Web application should show the place that crime happen places on the map by using the coordinates.

F. Non-Functional Requirements.

- The Mobile application enables to access multiple users simultaneously.
- The system allows the data stored to be available for later analysis.
- The system did not allow unauthorized access to the web application.
- The system protects the information of the users.
- The mobile application can be accessed anytime and anyplace which has internet connection.

IV. EVALUATION AND RESULTS

In this, we describe evaluation of our approach and the developed system while evaluating the objectives achieved how the project deviated from its original specifications and the circumstance identified during the time period of the project. This will give the idea of the measure that have been taken to handle the problem occurred and knowledge which have been gathered by supplying solutions for such issues.

Summative evaluation was used as the evaluation method to find how the system functions and whether it is up to the expected level to fulfil the clients' requirements. At the system finalizing stage this evaluation is done to evaluate the product's stability. In summative evaluation a prototype with most stable build is shown to the client and the feedback is taken to find how far the system is success. In here the using prototype must be very much alike to the final product's functions and features.

The overall Evaluation of the product was carried to verify whether the system's final outcome meets the functional requirements of the users and the successfulness of the system tasks and the functions of each component are also evaluated here. This was done by considering the functional requirements specified by the system specifications. The prototype was given to ten people and they were asked to rate the system based on the following attributes their responses have been recorded and summarized as below (Table 1).

Table 1. Summary on Used Technologies

Mobile App	Developed System
Efficiency	94%
Accuracy	78%
Cost Reduction	96%
User Friendliness	98%

Table 2. Summary on Used Technologies

Algorithm	Accuracy
Nearest policemen Identification	96%

According to the results, more than 95% of candidates have been satisfied by proposed system.

V. CONCLUSION

The results and outcomes generated in relative to the specificity of the problem domain are enlarged into wider concepts depending on logical assumptions. This chapter aims to clearly emphasize the outcomes and findings of the project and to determine way of these outcomes and findings can be matched in different contexts that are similar to the problems which are solved by the developed Emergency alert system for reporting crime issues to nearest police station. Furthermore, future enhancements for the developed System have suggested finding out ways to give in addition features to the system and using it outside the business subject in use.

Using this application, it happens a crime people can inform police immediately through informing current location to police. The message which generate this application basically send to nearest police mobile and nearest police station. In case of information confirm police, officers can arrive crime location in correct, manners in correct time. Basically, this application will help up to certain point reduce the crimes in Sri Lanka. The initial aim of this project is support police to identify the exact places and aware the people about crimes. Generate a report regarding crime arias. Through this application possibility to reduce happening of crimes in Sri Lanka. The development team implemented this system to determine its ability to satisfy the entire functional and non-functional requirement with special qualities such as flexibility, reliability efficiency and etc., to overcome the drawbacks identified in the system. The study found out that it is feasible to use the language ASP.NET in C#, SQL Server 2012 as database and java in android studio used to develop the mobile application to develop the project.

It's is a mobile application and web-enabled project, so this mobile application offers user to install the application and enter data. This is very helpful for the user to enter the desired information through so much simplicity. The user is mainly more concerned about the validity of the data, whatever he is entering. In Web server admin provided the option of monitoring the records entered earlier. Data storage and retrieval will become faster and easier to maintain because data is stored in a systematic manner and in a database. Decision making process would be greatly enhanced because of faster processing of information since data collection from information available on computer takes much less time than manual system.

This system allows to get information about the crime issues in the relevant city. This gives efficient and cost effective. Mobile application can be access by defined user categories by verifying their username and telephone number and web server can only be access by the admin by verifying the username and password. Client machines can be Windows xp, Windows 7, Windows 9 or Windows 10. Server computer should have operating system Windows xp, Windows 7, Windows 9 or Windows 10 and should be installed Visual Studio 2012, SQL Server 2012 and tool set.

ACKNOWLEDGMENT

It would not have been conceivable without the generous support and help of numerous Individuals who were there throughout the project. I might want to expand my thanks of every one of them. First and foremost, I would like to express sincere gratitude to the supervisor of this project for all inspiration and direction to do this report and kind support and assistance given throughout the report writing. The courage he gave by assessing during our report preparation and providing quick response, are very much valued. And I would like to thank everybody who were bolster us for the achievement of our project.

REFERENCES

Ann Dryden Witte (1996) Research suggests that some social and criminal justice policies can affect the crime rate.1996.09.00

Cameron, Samuel.(1988) "The Economics of Crime Deterrence: A Survey of Theory and Evidence." Kyklos 41: 301–23.

Corman et al (2000) "A Time-Series Analysis of Crime,Deterrence, and Drug Abuse in New York City." American Economic Review 90 584–604.

Ginger Hartman (2012) Online Dating & Relationships,[online] Available at:<http://police.uw.edu/services/crimealertprogram/

Jonathan Klickjonathan Klick(2003), Changes in the terror alert level set by the Department of Homeland Security provide a shock to police presence in Washington.2003:90:90.

Ken Maguire (2006)"Crime alerts at your fingertips in Boston"2007.Boston 89.00.89.

LI YUAN et al (2007)"Wireless Messaging Used to Fight Crime." Netherland.87.890.99.

E-COMMERCE WEB APPLICATION FOR AGRICULTURAL DEVELOPMENT IN SRI LANKA

RL Jayakody[1], DMR Kulasekara, and ADAI Gunasekara
Faculty of Computing, General Sir John Kotelawala Defence University Rathmalana, Sri Lanka
[1]ravinajayakody@gmail.com

Abstract- This research has been conducted to develop an e-commerce web application based on agriculture, which can be used by cultivators, and consumers across the nation. This application will also include the capability of adjusting and functioning on mobile phones.

This e-commerce application acts as a mediating platform between the cultivators and the consumers making it possible for both parties to directly interact with each other without the aid of a third-party, who in Sri Lankan scenario gains the highest profit by getting the agricultural products from the cultivator to the consumer. Due to this web application, users can also be aware regarding the agricultural products in other geographical locations as well.

This e-commerce website has a great possibility in increasing profit for the cultivators in the rural areas. This web application could also be predicted to drastically lower the consumer end-price of agricultural goods which can encourage people to consume more healthy food. On the other hand, this could be a great incentive (morally and monitory) for the cultivators to cultivate and promote their geographical and agricultural products. Therefore, the country can develop more in the path of agriculture.

Keywords- e-commerce, web application, agriculture

I. INTRODUCTION

Currently most of the cultivators in rural areas are facing difficulties in selling their agricultural products to the consumers. This problem has arisen due to many reasons. One such reason is due to the third-party that is being involved in the process taken place from the cultivator to the end-consumer, since this is a long process and therefore it is very time consuming. For example, when a consumer requests for certain agricultural products from a third-party, the third-party will then have to contact the cultivators to receive the expected products, hence having a third-party consumes time than the consumer contacting the cultivator directly. Involving a third-party extremely increases the end-prices of the products according to the Sri Lankan scenario. This can in turn lead to less consumer consumption of the desired products from cultivators and consume other alternatives. As a result, it will occur in cultivators' mental and lifestyle degradation. Even though there are a vast variety of agricultural products that is being produced, they are not exposed to places in Sri Lanka other than the geographical regions they are produced in, thus lack of awareness among consumers results in wastage of agricultural products.

This agriculture aiding e-commerce application using the web platform is developed as a solution for the above-mentioned problem. All the users should be registered members of this application in order to access the web application. It is possible for the users to display the text in the interface in either Sinhala or English. They can also upload images regarding their products. Once the image is uploaded the application has the capability of identifying the content of the image to recognise which product type it belongs to. The location of the cultivator can be either uploaded manually or automatically detected via the application. The input public information about the cultivators will be displayed for the consumers. The consumers can select or search for their expected agricultural products. A list of cultivators available with the desired product will be displayed. Along with those information, the locations of those cultivator will be displayed to the consumer. The nearest consumer with the available product will displayed at the top of the list. The consumer can place an order to the cultivator regarding the need of a certain product(s). When an order is placed the cultivator will get notified. The delivery of the agricultural products will then be in process after it is agreed either to pay via PayHere or cash on delivery.

II. RELATED WORKS

A. Impact on Market and Retailers

This research focuses on the cultivators in rural areas, agricultural economy of Sri Lanka and the preferences to propose the best-suitable e-commerce web application for the user island wide. It could be seen that this could be achieved at a software level. Implementing an e-commerce web application which is successful encompasses the knowledge and understanding of the limitations and diminishing the negative impact whilst simultaneously increasing the benefits of e-commerce businesses. Economists have made speculations that the price competition can be strengthened by e-commerce as it escalates the capability for the consumers to gain information about products and its prices. (Shahriari et al., 2015) Transforming of social and economic organisations by the vast technological advances which as a result brings millions of people in the entire world together. (Ignacio et al., 2017) Therefore, in this modern and more competitive environment with all the technological advances, medium-sized and also small businesses ought to be prepared and cognizant of the globalisation which is being experienced and cooperate with the electronic environment to acquire competitive advantages. (Ignacio et al., 2017) Electronic commerce affects the supply and demand of goods and service in different businesses, the structure of the business and competition in several sectors. The spread of e-commerce will continue in reducing the costs of transaction and also the costs of production and facilitate integration of new companies into the markets and escalation of competition will be anticipated in the short run. Thus, lower prices, increase the levels of quality and create novel and more products in a wide-range consequently increasing the growth of economy and welfare. (Ignacio et al., 2017)

B. Language Literacy and Computer Literacy

For e-commerce, computer literacy data can provide immeasurably to an understanding of the demand and supply of skills in the global, knowledge-based on economy.

Table 1. Thecomputer literacy rate by sector and province in Sri Lanka for the years 2016 & 2017 during its 1st 6 months.

Sector/Province	Computer literacy rate (%)	
	2016	2017
Sri Lanka	**27.8**	**28.3**
Sector		
Urban	38.5	41.1
Rural	26.1	26.5
Estate	9.9	9.5
Province		
Western	38.5	38.6
Central	26.0	30.2
Southern	27.2	29.1
Northern	19.9	15.1
Eastern	13.4	13.7
North Western	27.3	28.3
North Central	21.6	20.9
Uva	18.5	15.5
Sabaragamuwa	23.4	26.8

According to the "Computer Literacy Statistics – 2017 (First six months), from the Department of Census and Statistics, Sri Lanka" the results for computer literacy in various age categories, sectors and provinces, language literacy fields, and as well as in various occupational groups are shown in the tables as displayed above.

A person in the age limits within 5 to 69 is considered as a literate in the field of computing if he/she is capable of using a computer on his/her own. For example, if a child who is 5 years old has the capability of playing a computer game then, he/she can be considered to be a literate in the field of computing.

In the first half of 2017 for Sri Lanka, the general rate for Computer Literacy is reported as 28.3%. The above survey results demonstrate an increase of 0.8 percentage points from the first half of 2016 to the first half of 2017. The highest literacy rate in computing is shown by the Urban sector with a percentage of 41.1 amongst the residential sectors, whilst the rate of computer literacy for the Rural and Estate Sectors are shown as 26.5% and 9.5% respectively. When taking the provinces into consideration, computer literacy with the highest level is

stated from the Western province with a percentage of 38.6 whilst the computer literacy with the lowest percentage is reported from the Eastern province with a percentage of 13.7. (Computer Literacy Statistics, 2017)

Table 2. Computer literacy rate amongst computer aware employed population within the age limits between 15 to 69 years by Occupation during the first 6 months of the year 2016

Occupation group	Computer literacy (%)
Sri Lanka	**64.8**
Managers, Senior Officials and Legislators	76.1
Professionals	90.7
Technicians and Associate Professionals	88.7
Clerks and Clerical support workers	91.5
Services and Sales workers	69.1
Skilled Agricultural, Forestry and Fishery workers	22.3
Craft and Related Trades workers	43.4
Plant and Machine operators and Assemblers	53.5
Elementary occupations	27.1
Armed Forces Occupations & unidentified occupations	86.6

The table displayed above has discovered that Skilled Agricultural workers as well have the least knowledge in the field of computer literacy, in Sri Lanka. (Computer Literacy Statistics, 2016)

Table 3. Computer literacy rate by Language literacy for the first 6 months in the year 2016

Language literacy	Computer literacy rate (%)
Sri Lanka	
By Language literacy (age 10 - 69)	
Sinhala	33.5
Tamil	26.2
English	72.5

The above shown table reveals that the computer literacy is higher among those who are also literates in the English language. According to the survey it is displayed that the English literates have a percentage of 72.5 in computer literacy than the Sinhala and Tamil literates who has their computer literacy as 33.5% and 26.5% respectively. (Computer Literacy Statistics, 2016)

C. Existing Websites

There are a numerous number of websites in the world. As we have researched and analysed throughout various types of websites, I have come across many websites which are quite complicated to be used by people who are Non-English literates as well as non-computer literates. Certain websites which are currently implemented, such as European websites, Asian websites, Sri Lankan websites are taken into consideration and are listed below;

i. Amazon

ii. Royal Flora Holland

iii. ikman

The interface of amazon.com and Royal Flora Holland seems user-friendly almost for any person who even has a moderate level of knowledge in the English Language. Whereas for non-English literates in rural areas, such interfaces appear to be complicated and hard to understand since language translation into Sinhala is not available in these above-mentioned sites.

Figure 1. Graphical User Interface of an e-commerce website
Source: amazon.com

The graphical user interfaces on ikman however gives the impression of complexity for cultivators in rural areas even though there is language translation to Sinhala as well.

Figure 2. Graphical User Interface of an e-commerce website
Source: ikman.lk

The cultivators in rural areas in Sri Lanka expect their work to be done with much ease with a minimum number of clicks in the website due to the lack of knowledge and skills they possess, else the website is found to be complicated and mystifying and will eventually be rejected by them by tending to not use such websites for their purpose. They also require their tasks to be done in a stepwise manner one after another, which gives them a clear idea of how the process may take place. Consequently, the consumers outside the geographical area will be unable to know about the availability of the products by the farmers in that specific location and will therefore result in wastage of food.

All the above-mentioned websites give the opportunity for the users to allow photos of the products to be uploaded, nonetheless, none of these websites has the capability of automatically identifying the type of product which has been uploaded and retrieving the necessary data regarding the product instead of allowing the user to enter all the details in to the system to post and display about their product. Absence of such availability in current websites have been overcome by our system by using existing technologies in order to create such a website which will only entail a minimum amount of inputs to the system and a maximised level of output to the user. Thus, permitting the possibility of achieving the expected work done with a limited number of clicks.

D. Location Identification

In today's world, due to the advances in technology to make it much more convenient to the user in every aspect there are several researches being taken place daily around the world. Researches are being done based on tracking the location of a person and showing the route to a person from source to destination. Using of these services will guide the user throughout the route also notify them about the nearby places in which there's availability of goods and services according to their preferences. (Indunil et al., 2017) Google APIs are used for better context data for navigation purpose which is considered by more than a few researches. (A. Kushwaha, V. Kushwaha., 2011) Using the APIs for location in the website it makes the possibility to provide the users with much more convenience during the search for food products.

E. Food Wastage

The wastage of food worldwide is about third of all food which is wasted each year. This further indicates that the loss of food also means as the loss of nutrition which is occurring due to the worsening quality or degrade of nutritious crops. Food which are rich in nutrients such as fruits and vegetables have the rate of wastage which are higher than any other food product. Although there is only an inadequate amount of information about the loss in micronutrients in food value chains, an occurrence in the loss of Vitamin A from the loss and wastage of food has been assessed via studies. In view of the insufficient micronutrients globally, the loss of nutrients could have turn resulted in momentous impacts on the efforts that are taken to lower the amount of unseen hunger and undernutrition. ("Reducing food loss is key to end hunger and undernutrition by 2025 | IFPRI," n.d.)

Figure 3. Food Wastage
Source: ifpri.org

Food losses in developing countries such as Sri Lanka have substantial consequences regarding the income of cultivators in rural areas, who leads the cultivation of various food items and also them being the major percentage of the poor and undernourished populaces. The extent of the crops to be sold reduces when on-farm losses occur and will therefore lower the cultivator's income mainly, in rural areas. ("Reducing food loss is key to end hunger and undernutrition by 2025 | IFPRI," n.d.)

On April 29th, 2018, the Vesak full moon poya day, a pumpkin dansel was held by one of the farmers from Makulugaswewa in Galewela. It is also said by him, that he does not harvest any pumpkins as there are no market for them. (lakmali, n.d.) This could be the reason why the economy in Sri Lanka is descending as there is lack of support for the farmers in rural areas and that they are unable of selling their widely cultivated crops.

Figure 4. A pumpkin dansel held by a dissatisfied farmer Source: srilankamirror.com

In order to improve the security of food, increase the amount of nutrition, and income to cultivators in rural areas it is vital to reduce the losses in food. Nevertheless, particularly in developing countries, the extent of food losses where exactly in the value chain they befall for different products and countries is however indistinct, hence designing of targeted policies and programs to reduce the loss of food is strenuous. ("Reducing food loss is key to end hunger and undernutrition by 2025 | IFPRI," n.d.)

III. MATERIALS AND METHOD

During the stage of development of the e-commerce web application, HTML, PHP, CSS, JavaScript and MATLAB language are used as the required programming languages. The database used for this application is MySQL. The web application languages mentioned above use the software JetBrains PHPStorm, XAMPP, and MATLAB for the implementation of the website. Bootstrap is used as the front-end framework for the designing of the website interfaces. JetBrains PHPStorm software is used for the coding of the web application. MATLAB software is used for image processing using the MATLAB language. A web server is crucial to host the web application for which Amazon Web Services will be used. Google Maps and Geolocation APIs are used to track the location of the users. Google Translate API is used to translate the displayed language on the application either form English to Sinhala or vice versa.

This application is using a minimum number of requirements from the users, regardless of the fact if the user is a literate in the field of computing or not. The website is also using user-friendly interfaces, and the language displayed will be as per the users' preference since people in rural areas have a great tendency to have much of a lesser knowledge in any foreign language, such as English. Therefore, Sinhala or Tamil could be chosen in replacement. Both consumers and cultivators can register to this application as a common user and gain the capabilities of both buying and selling. This system has the capability of tracking the location of the cultivators and consumers, to provide them with best facilities in order to get the products transported to the consumer from the nearest cultivator possible. The location will be notified via geolocation. Image processing is used to recognise the products from the images uploaded by the user, to ease the use of this system to its users. This application is developed with the compatibility for both desktop applications and mobile applications such as smart phones.

The Figure 5 below shows the layered design architecture for the proposed e-commerce web application.

Figure 5. Design architecture of the proposed system

In order to develop an efficient and effective e-commerce web application, certain functions that must be fulfilled to manage all the processes in this application.

- Capability of providing secure authentication.
- Ability to input data by the users.
- Ability to store and maintain a database.
- Ability to manage the content in the website.
- Fast and accurate search function filtered according to the consumer's location.
- Prioritizing the contents displayed in the items page according to the user preferences.

- Ability to provide promotions and discounts.
- Use of report generation mechanisms.
- The "Help" facility to be always available for the users.
- Integrating of Email marketing.
- Including a wish list in the application to enable users bookmark items which they might prefer to purchase in future.
- A checkout which is easy to use.
- Various methods of payment handling such as; Cash on Delivery and PayHere.
- Should be a mobile friendly website.

The Figure 6 below shows all the activities that are done by the administrator and the users (consumer and cultivator) in the proposed web application.

IV. DISCUSSION

In Sri Lanka, at various locations, though people don't cultivate in massive quantities, they cultivate for a certain extent. These cultivators come across a huge problem since they are unable to sell their agricultural products to consumers in other areas in the nation. To solve this existing problem, this application allows all cultivators and consumers in all geographical locations in the entire nation to interact with each other. There are agricultural products that are unique to specific geographical areas, yet in which consumers in other areas have less knowledge on those unique products. Hence, consumers having less awareness also leads to wastage of products and cultivators quitting in producing them. To overcome this situation this web application allows every consumer to increase their awareness in all types of products displaying it via the web application. At most circumstances, there's always a third-party involved in between the consumer and the cultivator. Most consumers buy their products through the third-party for high prices. The third-party earns the most profit in the process from the cultivator to the end-consumer. Involving a third-party expects consumers to purchase products for greater prices than they could be consumed for lower prices when directly purchasing from the cultivator. The third-party demanding a high price could be a reason possibly due to the cause of transportation expenses from the cultivator end to the third-party since the products may be transported from far

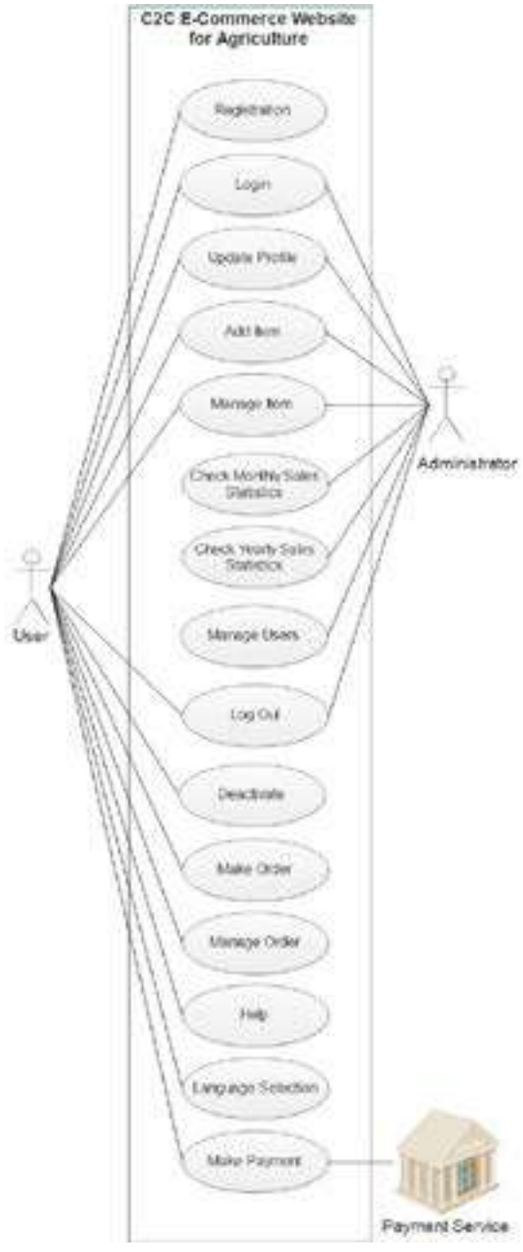


Figure 6. High level use case diagram of the proposed system

away locations as well. As mentioned before, the wastage of products due to the lack of awareness of consumers, another fact can probably be because the wastage of foods due to the long rides while the transportation is

being taken place. Thus, to avoid unnecessary expenses, direct interaction among the consumer and the cultivator is made possible through this application and also as a solution to refrain transportation from faraway locations to decrease the expense as well as the wastage of agricultural products, the application detects the current location of the consumer and displays cultivators available with the expected products by the consumer, starting from the nearest possible location onwards. Therefore, it can prevent any extra transportation costs and also all unnecessary expenses arisen due to the involvement of a third-party in the process Furthermore, this method can broadly decrease the amount of time that is being consumed during the entire process.

V. CONCLUSION AND FURTHER WORKS

The main aim of this desktop and mobile web application is to enable the cultivators and consumers from poor to rich interact widely with each other Island wide by being at any geographical location and introduce and sell their products which are also unique to a specific geographical location by making awareness through the mobile and desktop web application of such products among the consumers from other geographical locations as well., in order to make the nation develop in the cultivation of agricultural products.

ACKNOWLEDGMENT

We thank all the past researchers who have conducted their research explored in this field of study and the anonymous referees for their immense support and their useful suggestions.

REFERENCES

Indunil, E.J.C., Gunasekara, V.S., Jayampathy, K.A.T.S., Jayasooriya, H.D.M.P., Wimalasooriya, A.K.G.P.K., Gunathilaka, T.M.A.U., 2017. The best preferred product location recommendation according to user context and the preferences, in: 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (icter). Presented at the 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (icter), pp. 1–6. Https://doi.org/10.1109/ICTER.2017.8257820

Lakmali, n.d. Disgruntled farmer holds a pumpkin Dansel (pics) [WWW Document]. URL https://srilankamirror.com/photo-story/8364-disgruntled-farmer-holds-a-pumpkin-dansel-pics (accessed 7.22.18).

Reducing food loss is key to end hunger and undernutrition by 2025 | IFPRI [WWW Document], n.d. URL http://www.ifpri.org/blog/reducing-food-loss-key-end-hunger-and-undernutrition-2025 (accessed 7.22.18).

Shahriari, S., Shahriari, M., Gheiji, S., 2015. E-commerce and it impacts on global trend and market. Int. J. Res. –Granthaalayah Knowl. Repos. 34.

Ignacio Ortiz Betancourt., María del Carmen Meza Téllez.,

Mariel Terrones Castro., 2017. E-Commerce as Mechanism for Enhancing Micro and Small Enterprises: The Case of a Port Conurbated Area in the Southeast of Mexico, Vol. 8 Nr. 1: May-Aug 2017, ISSN 2411-4073

Amit Kushwaha, Vineet Kushwaha., Location Based

Services using Android Mobile Operating System, Mar 2011, ISSN: 2231-1963

Computer Literacy Statistics – 2016 (First six months) Department of Census and Statistics Sri Lanka, January - June – 2016, ISSN 2012-6565

Computer Literacy Statistics – 2017 (First six months) Department of Census and Statistics Sri Lanka, January - June – 2017, ISSN 2012-6565

Amazon.com
https://www.amazon.com/

Royal FloraHolland.com
https://www.royalfloraholland.com/

ikman.lk
https://ikman.lk/

Abbreviations and specific symbols

API	- Application Program Interface
CSS	- Cascading Style Sheets
E-Commerce	- Electronic-Commerce
HTML	- Hyper Text Mark-up Language
MySQL	- My Structured Query Language
PHP	- Hypertext Preprocessor
XAMPP	- Cross-Platform, Apache, MariaDB, PHP and Perl

SAFE ACCIDENT ALERT SYSTEM FOR REPORTING ACCIDENTS TO NEAREST HOSPITAL AND POLICE STATION

GSA De Silva[1], SCM de S Sirisuriya and RMM Pradeep
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]*mails2santhushi@gmail.com*

Abstract- Road Accidents are fast becoming one of the major issue of deaths in Sri Lanka. In some cases, they are accounted for from exceptionally remote areas now and again there is not really any vehicular traffic on streets. There are numerous factors that reasons road traffic accidents include Drunken driving, Lack of knowledge about Road Rules and regulations, High Speed, use of mobile while driving etc. A mobile application for Analysis of the Traffic Problems and Safe Accident Alert System has been proposed to overcome from this matter. There are different types of Accident Alert Systems all around the world. But in Sri Lanka there is no proper computer-based Accident Alert System has been developed yet. This mobile application tracks the location of accident using GIS and SMS services. With the help of this system, the systems alert the nearest police stations and hospitals. And, analysis all the traffic issues and road signals. The main objective of the system is to provide help and need for the vehicle user and detects the accident if occurred and informs the respective authority through wireless technologies and find the vehicle where it is, in order to give treatment for injured people. Many lives could have been saved if the required attention was given at the time of need. Leading people to use mobile phone for many tasks of daily activities. And this situation is correlating the need of this kind of situations.

Keywords- GIS, Location tracking, Alert Message

I. INTRODUCTION

With the increase in traffic problems and accidents has become one of the major issues in Sri Lanka. According to statistics of the Sri Lanka Police, accidents have seen an increase in the recent past with the steady increase of vehicles on the country's roads. According to the Police, last year alone 2,600 fatal accidents took place resulting in 2,817 fatalities, and during the first six months of this year 1,365 fatal accidents were reported with 1,313 fatalities caused. According to Police statistics 19,916 accidents have taken place in 2016 during the same time up from the reported 18,774 accidents last year. In this project, decided to implement a mobile application to develop a computer-based accident alert system. Using Information & communication technology this problem can solve successfully and improve the environment friendliness. By using this "SAAS" (Safe Accident Alert System) mobile application anyone can inform the locations of that accident happens. Anyone can use this mobile application from anywhere by typing their name and their mobile number. And then he/she can track the locations and send it to the nearest police station and hospital. Then user will get a confirm message whether the tracked location and other information's (date and time, username, user mobile number) are send or not to the police station and the hospital.

In Sri Lanka there are many road accidents, in mainly the Colombo district. There are various reasons for these road accidents. Most of the time these accidents happen either because of recklessness of the driver or the pedestrians. (Figure 1) Sometimes they are reported from very remote areas at times there is hardly any vehicular traffic on roads. Most of the times we may not be able to find accident location because we don't know where accident will happen.

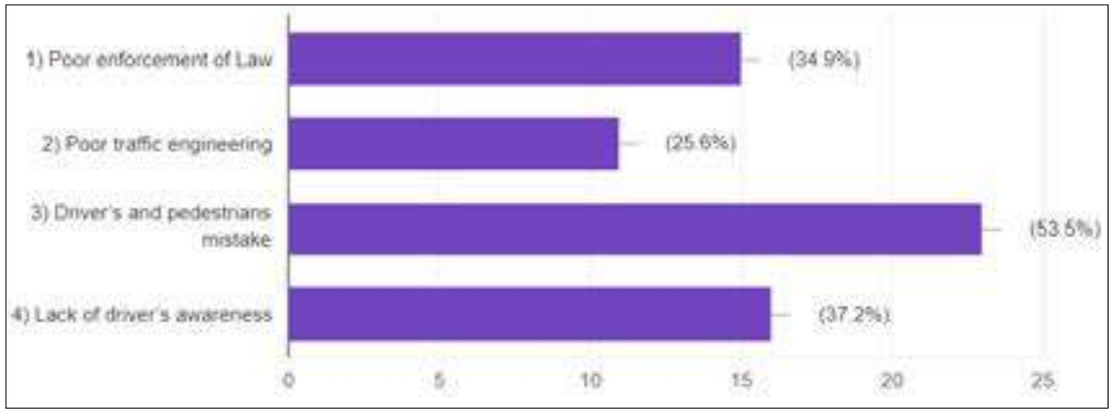


Figure 1. Bar Graph of most responsible issues for road accidents.

The purpose of the project is to find the vehicle where it is and give treatment for injured people, first we have to know where the accident occurred through location tracking and sending a message, and we have identified the following objectives.

- To critically study the existing computer-based solutions for, current issues in Traffic and Accidents in Sri Lanka.
- To design the system and develop the prototype properly.
- Testing and evaluation of the new Safe Accident Alert System.

II. LITERATURE REVIEW

Through the review, it will be thought about regarding the procedure and available practices in traffic problems and accidents.

Kaladevi, P. et al (2014) has developed a system for this issue by developing an Accident Detection System utilizing android smart phone from the accident zone. This framework has been created and executed utilizing the heart beat sensor based mobile technology integrated with the developing android smart phone. The application for accident detection which basically measures the heart beat rate utilizing heart beat sensor. In the wake of getting the flag from sensor this system sifts through the foundation commotion and recognize just the sound of the beat. At that point tally the time between each heartbeat to get the beat rate. In the event that there is any variety from the normal heart beat range, at that point the system identifies that might be a accident or not. At that point the system will instantly transmit the area of the accident to the pre-arranged contacts through Short Message Service (SMS).

Zhang, Dahai. (2005) has presented an analysis of the traffic problems and research on the traffic strategy in group urban development. Urban group development is aimed at avoiding blind urban extension. It is a system of ordered structure, supplementary function, optimised integration and shared construction that allows harmonisation of different city mode levels horizontally and vertically. However, extending the city group space leads to increased distances between cities. In this case, it is necessary to construct high-speed access and networks. The idea of an" integrated green traffic" strategy is a sustainable urban development traffic system with 'human priority' and is proposed for the present group city which is facing various traffic situations.

Obuhuma. (2012) has presented a Use of GPS with Road Mapping for Traffic Analysis. The application of GPS in traffic analysis is proving to be the best arrangement contrasted with other existing traffic management strategies like security cameras, human investigation, and speed governors and tachographs. Mapping of situational street activity speed at any given time draws out the coveted geographic examples and connections which are principal basic leadership instruments for traffic administration. The current convoluted traffic systems, activity speed and the gigantic number of the traffic members, requires advanced and programmed techniques for information

catch as the main best solutions to traffic control. For effective execution of activity administration and control measures, not exclusively should the specialized points of view of the measure be legitimately considered, yet in addition the mix with existing foundation and approaches.

Mukesh, P.R. (2010), has developed a vehicle system that are right now being used is some type of Automatic Vehicle Location (AVL), which is an idea for deciding the geographic area of a vehicle and transmitting this data to a remotely located server. To accomplish vehicle following continuously, an in-vehicle unit and a following server is utilized. The data is transmitted to a following server utilizing GSM/GPRS modem on GSM arrange by utilizing cell phone instant message or utilizing direct TCP/IP connection with following server through GPRS. The following server additionally has GSM/GPRS modem that gets vehicle area data by means of GSM system and stores this data in a database.

Kommineni, Rakesh. (2014) has presented Vehicle tracking and accident alert system. In this system, at first the GPS ceaselessly takes input information from the satellite and stores the scope and longitude esteems in AT89s52 microcontroller's buffer. If we need to track the vehicle, we have to make an impression on GSM gadget, by which it gets actuated. It additionally gets enacted by identifying accident on the shock sensor associated with vehicle. Parallelly deactivates GPS with the assistance of hand-off. When GSM gets initiated it takes the last received latitude and longitude positions values from the buffer and sends a message to the particular number or laptop which is predefined in the program. When message has been sent to the predefined gadget the GSM gets deactivated and GPS gets activated.

Hasan et al. (2009) used the GPRS service which made their system a low-cost tracking solution for localizing an object's position and status.

III. METHODOLOGY

A. Data Gathering

Qualitative and quantitative data required for designing the requirement specification for the new system were gathered by conducting a survey.

B. Approach

In the Safe Accident Alert System process first user should login to the Mobile application by providing user name and phone number to the system. Then the mobile application will track the coordinates (longitude and latitude) of the location where user is in (where the accident happens). Then those information; User name, User mobile Number, Date and time and coordinates of the tracked and will send to the web server that is located in nearest Police station and the nearest Hospital. (Figure 1 shows the design module of it) After taking the information from the mobile application those details will be displayed on those web servers. The location of the accident will be display on a map according to the coordinates that sent from the mobile application. Then the web application will send a notification for the user whether the information send successfully or not and give a compliment for the sender for his service Each information that users send will stores in SQL server database successfully and for the work with the application it required continuous internet connection.

C. Technology Adopted

It is necessary to use new technological methodology for the system. It is crucial to use acceptable tools to develop productive system. Use on any inappropriate tools can solely ends up in develop a system with unnecessary errors and faults and use of those badly chosen technologies

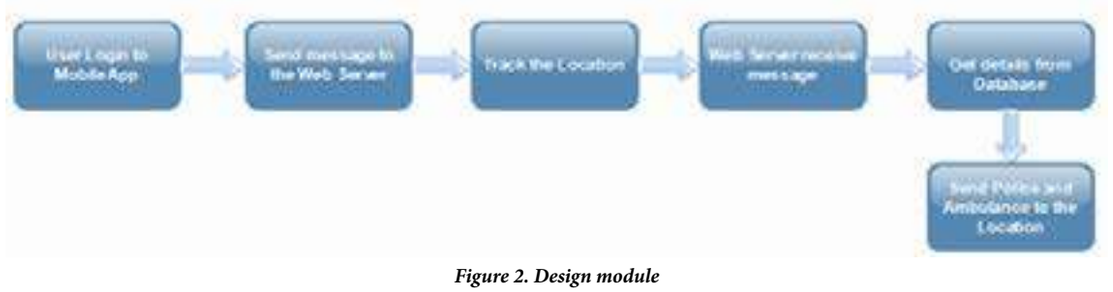


Figure 2. Design module

additionally can ends up in crashed when the new system implementation. Badly chosen technologies which can be extremely advanced and complicated will enable manufacturing a system with a top quality, however these technologies may result in develop a system that spend lots of time and resources to perform a task that is anticipated by the system. It is very important to use applicable programming language and the other necessary tools to develop a productive system. Therefore, these technologies and tools can help to develop the system among a minimum development time the most objective of developing this type of an application is to produce the users more efficient work system instead of doing manual approach. Because of that we should use the most applicable tools available in the market to develop the system. Technological considerations - followed during the development of the system Efficiency and Performance Re-usability and flexibility object-oriented development support so according to the Safe Accident Alert System java and android studio used to develop the mobile application. According to that requirement system has developed by using C# and using SQL database to run on windows operating system.

• Web Application

The programming language that is going to apply as the developing language for the system development turned into significantly trusted accuracy, performance. When considering all these technologies which can be associated with the Safe Accident Alert System. The proposed system can be applied a web-based technology. The .NET Framework comprises of the basic dialect runtime and the .NET Framework class library. The common language runtime is the establishment of the .NET Framework. You can think about the runtime as a specialist that oversees code at execution time, providing core services such as memory management, thread management, while also enforcing strict type safety and other forms of code accuracy that promote security and robustness. The class library is a thorough, protest situated gathering of reusable kinds that you can use to create applications extending from customary order line or graphical UI (GUI) applications to applications in light of the most recent advancements gave by ASP.NET, for example, Web Forms and XML Web administrations programs keep running on the .NET Framework, it keeps running on a virtual execution framework called the common language runtime (CLR) and a consolidated arrangement of class libraries. The CLR is the usage by Microsoft of the common language infrastructure (CLI) and it makes execution and improvement conditions in which languages and libraries cooperate with no imperfection. C# source code is accumulated into a middle of the road dialect that adjusts to the CLI determination. The IL code and assets, for example, bitmaps and strings are put away on circle in an executable record called a gathering. It contains a show that gives data about the get together, for example, types, adaptation, culture, and security necessities.

At the point when the C# program is executed the get together is stacked into the CLR. In view of the data in the show CLR may take different. At the point when the security necessities are met, the CLR performs just in time arrangement and change over the intermediate language code to local machine code. CLR additionally gives different administrations, for example, special case taking care of, and asset administration. The accompanying representation demonstrates the relationship of the normal dialect runtime and the class library to our applications and to the overall system. The illustration also shows how managed code operates within a larger architecture.

• Mobile Application

Android gives you the opportunity to put into impact your own device determinations and drivers. The hardware abstraction layer (Hal) presents a notable approach for creating programming hooks among the android stage stack and your hardware. The android working machine is in like manner open source, so you can make a commitment your own particular interfaces and enhancements. Android is an open source, Linux-based programming stack made for a wide array of devices and form factors. Android moreover comprises of a rigid of centre run time libraries that offer the vast majority of the ability of the java programming dialect, which incorporates some java 8 language functions that the java API framework uses.

• Database Selection

Consistent with the above eventualities most of the structures are used the square database to keep facts. It seems it is simple to control and perform. So, the database put in force on the server must able to supply efficiencies operations. Consequently, the proposed system decided on the Microsoft SQL server as server. SQL server is the

inspiration of Microsoft's data base platform, delivering challenge critical performance with in-reminiscence technology and quicker insights on any information, whether or not on-premises or in the cloud, and also Microsoft SQL Server is an application used to make computer databases for the Microsoft Windows group of servers working frameworks. Microsoft SQL Server gives a situation used to create databases that can be gotten to from workstations, the Internet, or other media as well. Database administration or DBMS, store client's information and empowers them to change the data into measurements. Those frameworks enable clients to make, supplant and remove actualities from their database.

A database is an established collection of information. Actualities allude to the attributes of people, things and exercises. Square server stores each measurement thing in its own one of a kind fields. In square server, the fields identified with a specific character, thing or event are packaged on the whole to shape a solitary finish unit of records, known as a report. Each record is comprised of some of fields. No two fields in a record will have the equivalent area name.

Throughout an SQL server database design project, the evaluation of your project wishes identifies all the fields or attributes of interest. If your commercial enterprise desires trade through the years, you outline any extra fields or alternate the classification of present fields.

D. System Architecture.

System architecture is divided into main three layers. They are Application Layer, Presentation Layer and the database Layer. (Figure 2.)

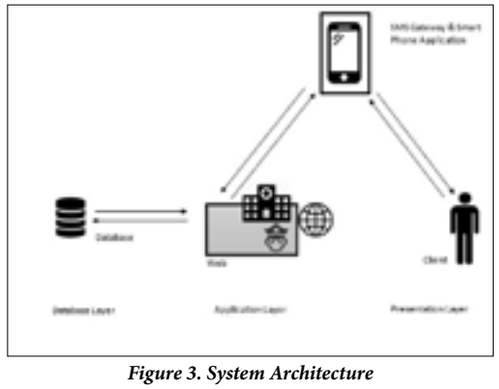


Figure 3. System Architecture

Data Layer: Data layer controls the data storage operations related processes of the overall Safe Accident Alert System. In this layer, SQL Server, has been used as the database management system. In this information layer SQL Server, has been utilized as the database administration framework. SQL Server is a light weight Database Management System that encourage with setting up information bases requirement for the product application Information will be put away in individual database which are assembled from client contributions through interfaces or by different procedures of the application layer, subsequent to controlling data at the application layer.

Application Layer: Application layer in charge of going about as the intermediate layer between presentation layer furthermore the information layer.

Presentation Layer: Presentation layer serves to information correspondence process by presenting to the users. It's give the interface to use to work with the product application. Data accumulated by presentation layer will be given to the application layer keeping in mind the end goal to control as per the given guidelines. As perspective in the presentation layer access the server and play out the tasks.

E. Modular Architecture

The discussed modules in the software architecture will be elaborated in this section to show how the modules will be dealing with the users' actions in each layer. The main modules and their sub modules are given below.

The Web application can be divide into five modules; Login and Authentication Module, Administrator Module, Confirmation Module, Map Module and the information Module

I. Login and Authentication Module

Only the authenticated users can access the system.

II. Administrator module

Administrator module is only can access by the admin. User cannot involve in this process. Admin need to be verified by the system and can change user names and

passwords of admin. Admin can get the details about users and details about locations.

III. Confirmation Module

In the confirmation module that in web application it will operate by the admin of the system. When the location of the accident will have identified, admin will send the confirmation message to the user that their process is successful.

IV. Map Module

This is the part of the web application. When user track the location and send the coordination; longitude and latitude of the location to the web server that exact location will display in a google map.

V. Information Module

This module gives the details about user who sent the accident location details. In this information module, it will record the name of the user, his/her mobile number, date and time and the location information.

The Mobile Application can be dividing into five modules; User Registration module, Message Sending Module, Tracking location Module, Confirmation Module and the information Module.

I. User Registration Module

In the mobile application, the person who uses the app should provide the user name and the mobile number of the user to track the location. All the registered member details will be shown by the system through the database interface.

II. Message Sending Module

User should be able to send message.

III. Tracking location Module

In this module, mobile application will track the location 's coordinates (longitude and latitude) And send it to the server. When the location is tracked location, coordinates will display in the tracking interface.

IV. Confirmation module

If those all information will successfully receive the web server, the user will receive a text message with a compliment.

V. Information Module

In this information module, the collected information such as username, user mobile number location details and location tracked date and time will be sent to the web server.

F. Software Design.

There are two user levels in the system. They are admin and ordinary people. Ordinary People will access the android application. (Figure 4) Admin will access the web application. (Figure 5)

Figure 4. Interfaces of mobile application

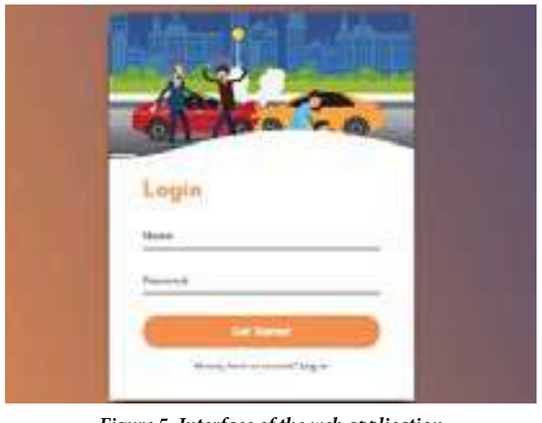


Figure 5. Interface of the web application

G. Functional Requirements.

- Mobile application should be able to track the location of the Accident.
- Mobile application should be able to send the particular details (user details, user mobile no, date, coordinates of accident)
- Web application should be able to get the information that sent from the mobile app.
- Web application should show the place that accident happen places on the map by using the coordinates.

H. Non-Functional Requirements.

- The Mobile application enables to access multiple users simultaneously.
- The system allows the data stored to be available for later analysis.
- The system did not allow unauthorized access to the web application.
- System protects the information of the users.
- Mobile application can be accessed anytime and anyplace which has internet connection.
- The designs will be created with appropriate themes, colours and font sizes. The mobile application need to be available all the time, and to everyone.

IV. EVALUATION

In this, we describe evaluation of our approach and the developed system while evaluating the objectives achieved how the project deviated from its original specifications and the circumstance identified during the time period of the project. This will give the idea of the measure that has been taken to handle the problem occurred and knowledge which have been gathered by supplying solutions for such issues.

Summative evaluation was used as the evaluation method to find how the system functions and whether it is up to the expected level to fulfil the clients' requirements. At the system finalizing stage this evaluation is done to evaluate the product's stability. In summative evaluation a prototype with most stable build is shown to the client and the feedback is taken to find how far the system is success. In here the using prototype must be very much alike to the final product's functions and features.

The overall Evaluation of the product was carried to verify whether the system's final outcome meets the functional requirements of the users and the successfulness of the system tasks and the functions of each component are also evaluated here. This was done by comparing the functions of the new system also considering the functional requirements specified by the system specifications.

V. RESULT

The prototype was given to ten people and they were asked to rate the system based on the following attributes their responses have been recorded and summarized as below (Table 1).

Table 1. Summary on Used Technologies

Mobile App	Developed System
User Friendliness	97%
Accuracy	80%
Availability	75%
Cost Reduction	98%
Efficiency	95%

According to the results, more than 95% of candidates have been satisfied with this system.

VI. CONCLUSION

The results and outcomes generated in relative to the specificity of the problem domain are enlarged into wider concepts depending on logical assumptions. In this aim to clearly emphasize the outcomes and findings of the project and to determine way of these outcomes and findings can be matched in different contexts that are similar to the problems which are solved by the developed Safe Accident Alert System. Furthermore, future enhancements for the developed System have suggested finding out ways to give

in addition features to the system and using it outside the business subject in use.

The main purpose of this project is to find vehicle where it is, most of the circumstances we will be unable to discover accident location since we don't know where accident will happen. To give treatment for injured people, first we must know where the accident occurred through location tracking and sending a message. The development team implemented this system to determine its ability to satisfy the entire functional and non-functional requirement with special qualities such as flexibility, reliability efficiency and etc., to overcome the drawbacks identified in the system. The study found out that it is feasible to use the language ASP.NET in C#, SQL Server 2012 as database and java in android studio used to develop the mobile application to develop the project.

It's is a mobile application and web-enabled project, so this mobile application offers user to install the application and enter data. This is extremely useful for the client to enter the desired data through so much simplicity. The user is principally more concerned about the legitimacy of the information, whatever he is entering. In Web server admin provided the option of monitoring the records entered earlier. Data storage and retrieval will become quicker and easier to maintain because data is stored in a systematic manner and in a database. Basic Decision-making process would be incredibly improved as a result of speedier preparing of data since information gathering from data available on computer takes considerably less time than manual system.

This system allows to get information about the Accident in the relevant city. This gives efficient and cost effective. Mobile application can be access by defined user categories by verifying their username and telephone number and web server can only be access by the admin by verifying the username and password. Client machines can be Windows xp, Windows 7, Windows 9 or Windows 10. Server computer should have operating system Windows xp, Windows 7, Windows 9 or Windows 10 and should be installed Visual Studio 2012, SQL Server 2012 and tool set.

ACKNOWLEDGMENT

I would like to take this opportunity to humbly extend my sincere gratitude and thanks to a few individuals who inspired and motivated me throughout the course of this project.

Firstly, my parents who have always been there when the times were both mild and though, providing us with all necessary guidance and inspiration. I would, have not been able to come so far if not for their blessings.

Special debt of gratitude is owed to the supervisor of this project for all inspiration and direction to do this project and kind support and assistance given throughout the report writing.

I also take this an opportunity to thank all the lecturers in the Department of Information Technology, who have extended their fullest support to me, in various forms.

To all others who have appreciated and criticized the work.

Finally, I believe I would fail in my duty if I forget all who have helped me in numerous ways. To all who I have missed to thank here.

REFERENCES

Bhumkar.S.P, Deotare.V.V, Babar R.V(2012)"Accident Avoidance and Detection on Highways", International Journal of Engineering Trends and Technology, Vol.3, Issue 2,pp. 247-252.

Hasan, Rahman K,et al (2009) 'Cost Effective GPS-GPRS Based Object Tracking System', Proceedings of the International MultiConference of Engineers and Computer Scientists 2009, Vol I IMECS 2009, 18 – 20.

John Whipple., William Arensman., Marian Starr Boler.(2009) "A Public Safety Application of GPS-Enabled Smartphones and the Android Operating System", Proceedings of the 2009 IEEE International Conference on Systems, pp.2059-2061.

Kardashyan, A. (2011) 'New Concept of the Urban and Inter-Urban Traffic', Proceedings of the World Congress on Engineering and Computer Science 2011, Vol II, WCECS 2011,19 – 21 October, San Francisco, USA.

Kommineni Rakesh. (2014)"Vehicle tracking and accident alert system"(ROLL NO :110EC0220)

KoteswaraRao.P., Haneef Babu.SK.MD.(2013) "Scening of an Accidental Alarm System of Vehicles With a Heartbeat Sensor", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Vol. 2, Issue 2, pp. 191-194.

Matsubara.A., Tanka.S.(2002) "Unconstrained and Noninvasive Measurement of Heartbeat and Respiration for Drivers Using a Strain" ,SICE 2002, pp.1067-1069.

Obuhuma, J. I. , Moturi, C. A. (2012)" Use of GPS With Road Mapping For Traffic Analysis."international journal of scientific & technology research volume 1, issue10, november 2012 issn 2277-8616.

Kaladevi,P. et al,(2014)"Accident Detection System using android smart phone from the accident zone"2014 -223-0090-00.

Mukesh,P.R. et al,(2010)"Real Time Web based Vehicle Tracking using GPS"World Academy of Science, Engineering and Technology 61 2010

Zhang,Dahai Jin Wenzhou. , Wang,Duanyi.(2005) " analysis of the traffic problems and research on the traffic strategy in group urban development".

A SOFTWARE SOLUTION FOR IMAGE IDENTIFICATION AND ARTISTIC SKILLS FOR VISUALLY IMPAIRED PEOPLE USING BRAILLE

ODS Fonseka[1], N Wedasinghe
Department of IT, General Sir John Kotelawala Defence University Rathmalana, Sri Lanka
[1]sasankafoska.dsf@gmail.com

Abstract- At present, visually impaired individuals have no method to perform or engage much in art. They do not draw paintings because of the absence of the ability to draw a world unknown and unseen before. New immerging software are developed within the world and the success of it is enjoyed only by the visually impaired who can afford it. The conversion of an image to braille and getting a braille printout of the image is the main aim of the new software application. By studying previously developed systems, new features were identified. Reducing the complexity of the software solution and providing the main output of a well converted image from the basic shape to complex image conversions will be made possible. With the results of the survey conducted for the research it emphasized the necessity of a software solution to give the opportunity to the visual impaired to engage in art using braille.

Keywords- Braille, Visually impaired, Art, Image Identification

I. INTRODUCTION

Visual impaired refers to when you lose part or all of your ability in vision. Life beyond an unseen world and a world to experience while feeling it is done by them. It is also referred as Blindness. Blindness is not a reason for failing in life and it is a challenge to make the outer environment in reach with other perceptions except vision. Perceptions refer to the ability to hear, feel by touch and smell to visualize the outer world. As stated below, " A loss or lack of vision does not deny an individual of all aesthetic pleasures since vision is only one of the senses which any of us perceive the world around us" (Irving Faber Lukoff, Oscar Cohen(1972)).

285 million people are estimated to be visually impaired worldwide, 39 million are blind and 246 have low vision according to the World Health Organization Reports (www.who.int, 2017). About 90% of the world's visually impaired live in low income settings (Deepti Samant Raja, 2016). With that scientists and doctors are concern to build new technological methods for the blind to retrieve their sight back. In reality, those methods are still in the experimental stage with new technological advancements.

In addition, those methods are normally considered as expensive technological equipment or surgeries which will be further explained in the literature review. As mentioned above the newly implanting methods will not be able to be afforded by the visually impaired people because there are expensive.

Due to the above reasons the necessity of a software solution for visually impaired people arise. In order to accomplish it the language of "Braille" was released by Louis Braille. He was completely blind and his touch sense characters of the alphabet helped the blind to read and gather knowledge about all areas in the unseen world. The alphabet was composed with embossed dots according to each specific letter. Braille is currently the most spread and successful methodology for blind people to interact with the world in aspects of learning and knowledge.

For visually impaired, blind visits or journeys around the world offer little or no information beyond verbal descriptions about the appearance, shape or feel of the actual pieces of art. The blind are at a serious disadvantage of experiencing the world beyond, due to the fact that they are not capable to touch all the objects in the world and some objects are impossible to touch with the human hand. The blind must rely upon the use of their remaining senses to develop and conceptualize a work of art and "touch, kinaesthetic experiences, and audition are the most important sensory avenues used for this purpose". They experience the world primarily through their sense of touch. A visually impaired individual "can gain knowledge of the spatial qualities of objects only by touch observations" and in order to do so "direct contact must be made with the object to be observed" (Lowenfeld, 1973).

Visually impaired people use their sense and memory to recognize where they are with the aid of a white cane. But the main concern of this project is the ability to identify the real image by touch and help the person to make his own image in the place he or she was told. "Creating art can provide positive feelings of accomplishment and achievement" for students with severe visual impairments (Shih and Chao, 2010). Therefore the conclusion on the participation of the visually impaired towards art can be stated out with a proper method to identify images around and visually recreate them in their mental perceptions.

The main objective of this research is to invent a more effective method for image identification and artistic skills using computerized braille. Researching on previous technologies used in computerized braille and in addition finding the new features to be included with the software to make it more effective to the visually impaired people is another main objective.

II. LITERATURE REVIEW

When considering the work done in the previous braille convertors and image identification softwares, all the systems will be reviewed and it will be beneficial to get the best features in order to develop the most effective software.

The problem existing in the world is the accuracy of whether image identification and artistic skills methods are beneficial for visually impaired in order to make themselves also a part with awareness on the environment and gain creativity through it.

Considering image identification there are many methods to identify images and the most common method used at present is imagination through sense perceptions which are not accurate to get a clear image of all the objects in the world. Using touch and white canes to identify the places and objects around with smell and taste senses to get awareness about the around environment also can be found commonly. The current techniques used for processing an image are image acquisition, image enhancement, filtering and edge detection.

Template matching with rule base is one technology used processing an image. The process consists of 5 main steps. Conversion to binary is done with these 5 steps such as binarization and pixel inversion, noise removal, segmentation and clustering, line identification and character extraction. (Mohd Solahuddin Bin Jaafar ,2011). In order to identify the processed image and braille characters which will be observed a GUT is introduced. The accuracy will be ensured by this GUT. The ability in displaying braille characters will be complied with a hardware device using Led. Though the system the visually impaired will be able to easily get a text picture with the hardware device to input a digital image to the software. A camera can be used in this instance. After that the input will be converted into braille characters.

In another previously invented converter the program involved in the recognition process was prepared with paint and imported into the OCR algorithm (Mohd Solahuddin Bin Jaafar , 2011). In the program the user can import their text images using scanners, digital camera or they can make it with Paint. The output such as text document can be printed out or observed on the computer screen. The initial setup of the project with the use of scanner of smart phone for image digitization, personal computer for image processing and printer for output observation.

Accurate text details provision can is achieved by a system called the Arabic Optical Braille Recognition System Technology. In this system the ability for the blind people to read text documents with more detail is given a solution with Optical Braille Recognition (OBR). The main hardware devices are the flatbed scanners and OBR software (Mohd Solahuddin Bin Jaafar, 2011). The braille letters are identified and converted into string characters. The braille characters are divided into cells and each character consisting in the cell is converted to text. It is a character recognition method used not involving any image.

New methods to improve semantic memory from the task are introduced with improved technologies by examined picture matching (Heller, 2002). The visually impaired were given to match images from what they felt. Four images were touched and felt by them and to test the accuracy of the matching it was timed. Within the people there were blindfolded people, people who were late blind, low vision and congenitally blind people. The very low vision people are braille users and used white canes for mobility. They tended to refer to themselves as blind, since they had no remaining pattern vision, or it was minimal. Most of these persons claimed that they could not see the close hand motion.

This system is to recognize characters for a single braille document and present an extensive review for braille. Recognition systems and related research efforts. This Braille recognition system is flexible for the size of the scanned image. It is improved in each step starting from the image acquisition until the Braille cell recognition final stage. The system includes an image acquisition stage, image pre-processing for noise removal, feature extraction, modified image segmentation, feature extraction, and then character recognition. In addition the system is applicable to any language and to both Grades one and two.

In the system made by Bhagya R Navada, Santhosh K V, Prajwal S and Harikishan B Shetty of Dept. of instrumentation and control engineering , Manipal Institute of Technology ,Manipal, India it is a system to detect colour for colour blind and identifying the edges of an image. Because of colour blindness the exclusion from some job opportunities can be found. The main reason is losing the ability to distinguish between colours. A solution for this problem is given by this new technology. To identify the colours image processing is used by representing the edges of images in similar colours. Colour detection and edge detection is done on a LabVIEW platform. By using a wireless camera the images are captured with its original colours and processed with LabVIEW colours. The processed image will be represented on the panel of the LabVIEW panel.

III. METHODOLOGY

In order to implement such a software first the image should be identified and filtered according to the following specifications. Removing the image noise, Sharpen edges, filter main image components and edge detection. The conversion to braille can be done after analysing the images edges by assigning a braille letter along the edge.

This methodology will be more effective than printing a typed braille image on a paper through the computer, minimizing the wastage of time with improved performance.

The research was mainly done in order to develop the artistic skills of a visually impaired based on the below 2 criteria's.

1. Image identification skills
2. Artistic development

After studying the different systems used, the pros and cons are identified as follows. Every system was able to do a conversion as they expected, but the final outcome is not very effective in comparison because at present a visually impaired individual can be given a much better software for image identification and continue in art through it.

First the image can be embossed by braille letters with the outline of the images to identify what this object. A feature to identify objects such as large monuments and structures with the conversion to braille will be achievable. Image colour identification to paint colours and description of the image on what image that the visually impaired is going to identify (Description translation input to braille) can be included. Texture identification through braille on the image for light and dark colours will able the artist to improve drawing skills.

Consisting of a foundation drawing converted to braille for beginners, colour identification braille conversion to draw art using colours, displaying the names of the colours of the original image to the braille converted image to draw and paint the image, colour intensity drawings for advance shading after practise from the initial stage and braille converted image which doesn't obstruct the image after converted to draw are considered.

The software solution consists of the following modules in order to get the final output. As inputs the JPG or png format images will be uploaded into the system for conversions in the image acquisition module.

Pre-processing is done for Image enhancement and feature extraction. Through these steps the image will be more elaborated to a format for the edge detection for the braille conversion. Converting to grayscale, removal of noise and image convolution will be done in the image pre-processing module.

In the image segmentation module background removal and binary thresholding will be applied. A Sobel filter is used to reconstruct the border of the main object in the image. Next, threshold to Zero, Inverted thresholding is performed to the image obtained from pre-processing. The image is then subjected to a combination of binary thresholding and Otsu's thresholding and a binary image of the image is produced.

The Edge detected image will be converted to braille by creating an image canvas and further more converting the ASCII image into a braille converted image. The creation of the braille image is mainly done by the ASCII art creation and image data to braille dot map algorithm.

Main steps of proposed new software solution are as follows; Adding new artist details, Consisting of a grid view ,editing or deleting details ,uploading images for further conversions ,braille conversions according to the skill level ,In this step, image processing steps; image pre-processing, noise removal, edge detection of overlapping and isolated cells, feature extraction and conversion will take place ,viewing the converted image and printing the converted image.

The image can be embossed with braille letters first with the outline of the images to identify what this object is. The embossed outline for the visually impaired to identify images will be according to the skill levels as in figure 1.

The design of the software solution is elaborated from figure 1 on all the modules of the application in order to get the final output braille images with additional features.

In the design the details of the artist will be stored and the images of the conversion will be saved within the database for further drawing requirements with more usability of the software. Image conversions will be done under the required stage of art for the blind artist. Three main stages such as Amateur, Intermediate and Professional will be

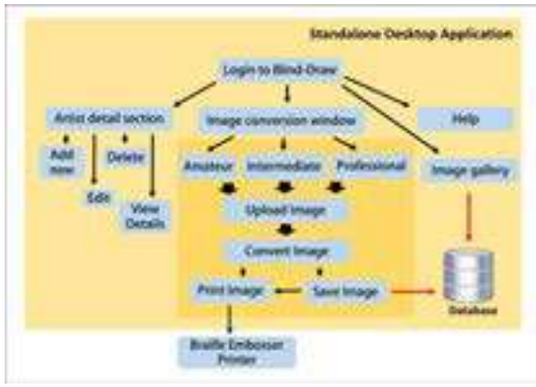


Figure 1. System Model Design
Source: Author

provided to start by drawing and improve the skill into the next level.

Simple navigations to access to features are used in order to make the software more user friendly and easy to access. The software solution is made from a ribbon form with user controls designed by C# and DevExpress for making the software more efficient to use.

IV. RESULTS

Considering the edge of the image the conversion should be done for the blind artist first by identifying the image in braille and there after use it as a basic foundation drawing to draw it. The printed braille output should be generated with the standard braille letters and though that the artist can engage in art.

A blind individual is not capable of using complex images as a beginning to drawings because an unknown world and it will be hard to be recognized with braille. Simple shapes such as circles, triangles, squares rectangles, trapezium will be converted and practised to draw via the hardcopy output of the software application.

In the conversion of the image primary images were selected at first. The basic shape of a circle as denoted in figure 2 will be converted into braille for the user to identify the shape of it. There after the user can use a braille conversion to colour on it with the respective colour denotations with the converted image.

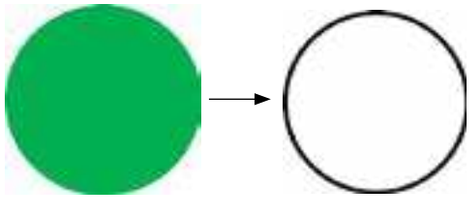


Figure 2. Circle image and edge detected of the circle
Source: Author

The edge of the JPG image will be extracted after pre-processing techniques using the image conversion module and the edge will be inverted for the braille conversion of the input.

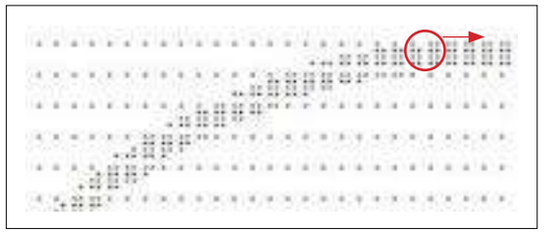


Figure 3. Enlarged portion of a braille converted image
Source: Author

The braille converted image will be converted as in a format which the blind people read normal braille as sentences and there after the artist should follow the image conversion to identify the image under the instructions provided with the braille conversion. The artist is guided with a starting point at the first line braille letter point which will be provided in the instructions.

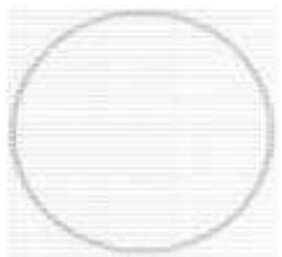


Figure 4. Braille converted image
Source: Author

In figure 4 it represents the braille converted image. The edge detected image will be converted to an ASCII image canvas and plotted with the braille dots within the conversion. Standard braille letter translations are used in order to make the blind artist to take his or her hand along the conversion to identify the image clearly. The image will be enhanced to the user to identify the image with the description the image as shown in figure 5.

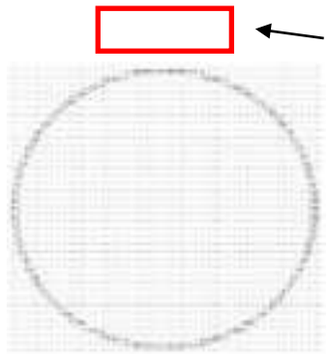


Figure 5. Braille converted image with the description
Source: Author

The blind artist can use a pencil or other sketch medium to draw on the printed copy of the conversion and there after using memory to remember the shape in the mind they can use it for further drawings.

Adding more skill the blind artist can convert the image according to the shape and colour in order to paint the image along with the braille printed hardcopy provided. Through this feature the user will be able to paint the respective shapes colour and in more improved skill level it can be used to indicate specific colours of specific shapes. An example conversion of the figure 1 of a green circle will be converted to denote its colour as shown in figure 6. In the conversion the colour will be represented in braille letters after translating to braille representation for colour identification and to paint that using that colour.

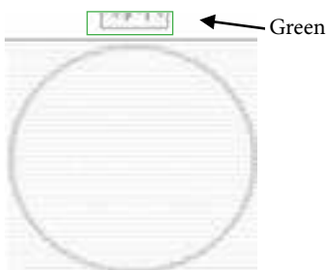


Figure 6. Braille converted image with the colour description
Source: Author

The colour identification and representation will be further improved for professional artists by indicating combined colours in one single hardcopy of a braille conversion. The primary stage of painting will be done by introducing the single colour with a primary shape in the amateur level of the application.

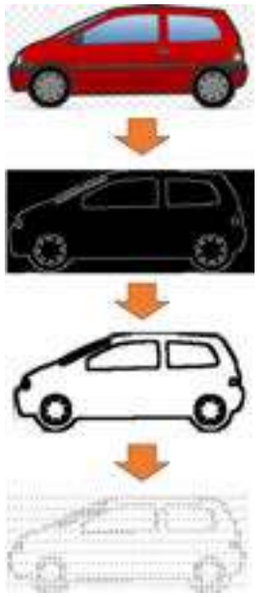


Figure 7. Braille conversion process of a complex image
Source: Author

The whole process for the braille conversion is shown in figure 7. A complex image of a car is represented in the diagram and such an image consists of smaller skill level shapes (circle, rectangles, trapezium etc.) which will be accessed in the amateur level within the application. The instructions for the blind artist should be provided in order to visualize the image in the mind and sketch it in the braille printed hardcopy.

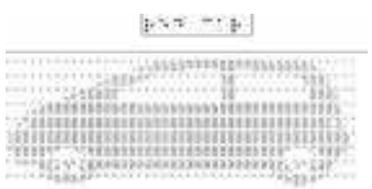


Figure 8. Braille conversion with description of a complex image
Source: Author

The description for painting the braille hardcopy is provided as shown in figure 8. The heading "Red car" is provided in the braille text line at the beginning.

V. DISCUSSION

The aim of this research was to identify the software solution effectiveness for Image Identification and Artistic skills for visually impaired people using braille and develop an application for blind individuals to engage in art. In order to create a fully functional image identifying computer based software through braille it is necessary to research on all the available technologies how beneficial it will for all the visually impaired personals.

The software was tested with the blind community of a sample of 20 students at the blind school and the students were able to identify the main standard shapes with the respective named label. Complex images were at a stage for the most creative students to identify and needed further improvements.

To recognize the effectiveness of such a research interaction with fellow artists, visual impaired people, lecturers and others who are concerned in such research areas was done from different age groups.

With the information gathered through the survey conducted the public awareness for the issue was gained on whether such an implementation in future will be beneficial for the visually impaired community. Considering the cost and software effectiveness the response to the braille converted image for image identification and art was highly appreciated. With less concern of people for such an area because they determine that the blind and visually impaired do not require any such subject modules a factor raised for doing a research whether creativity can be developed in each visually impaired individual by improving their image identification ability and artistic capabilities.

VI. CONCLUSION AND FURTHER WORKS

Image Identification and improving Artistic skills for visually impaired can be achieved by developing a computer based software solution. If such computerized technology is used, it will be a solution for the problem. The effectiveness of a software solution based on braille

will be simpler and less costly also will be influential for a successful mechanism to the visually impaired to identify the world through image identification and improve artistic skills through it.

To gain a conclusion on such a statement the information on the effectiveness of the software solution was gathered by a survey and the results gave the points to be considered in developing such a software in future.

As further work building an improved braille converting system with new features such as a method inventing a feature to guide the artist along the drawing and more improved colour representations within a conversion can be done. The using of braille conversion more mobile platforms will be also noted within future work.

ACKNOWLEDGMENT

I like to thank Mr. Dilum Chaminda who gave me the knowledge on image identification skills for the blind, all the individuals who participated in my surveys, all who were very helpful and provided feedback to complete this work.

REFFERENCES

Xuan Zhang, Cesar Ortega-sanchez (2003). "A system for fast text-to braille translation based on FPGAS"

Irving Faber Lukoff, Oscar Cohen (1972). "Attitudes toward blind persons"The American Foundation for the Blind.

Wolflang A. Slaby (1990). "Computerized Braille translation" vol.13, issue n2

Lauri Lydy Reidmiller (2003)."Art for the visually impaired and blind"

Minal Nerkar, Shweta Samangadkar, Kajal Lohar (2015). "Image Identification in Android Applications aiding visually impaired users", Volume 4 Issue 4

Kaustubh Bawdekar, Ankit Kumar and Rajkrishna Das (2016). "Text to braille converter", Volume 7, Issue 4

Gonzales R.C and Woods R.E (2008), "Digital Image Processing".

Mohd solahuddin bin jaafar (2011). "Braille conversion using template matching with rule base"

Morton A. Heller, Melissa McCarthy and Ashley Clark (2005). "Pattern Perception and Pictures for the Blind"

Bernard Gosselin, "From Picture to Speech Innovative Application for Embedded Environment".

Aisha Mousa, Hazem Hiary, Raja Alomari, and Loai Alnemer (2016). Smart Braille System Recognizer, Vol. 10, Issue 6

www.who.int (2017). WHO Media center. Fact sheet on Vision impairment and blindness.

Shih and Chao (2010)."Ink and wash painting for children with visual impairment" Volume: 28 issue: 2, page(s): 157-163

Deepti Samant Raja (2016). Bringing the Disability Divide through Digital Technologies

Lowenfeld (1973). Journal of Clinical Child Psychology Volume 3, 1974 - Issue 2.

Heller (2002) Journal of the Association for Research in

Otolaryngology : JARO 3(4): 488-498 (Journal)

A REVIEW ON DATA MINING TECHNIQUES TO PREDICT THE STUDENT PERFORMANCE AND DECISION MAKING IN EDUCATIONAL INSTITUTIONS

KG Abeywickrama[1], WJ Samaraweera, and CP Waduge
Department of Information Technology, Faculty of Computing,
General Sir John Kotelawala Defence University, Ratmalana, SriLanka.
[1]kgabeywickrama@gmail.com

Abstract- Education is significant as it represents the future of a nation. Most of the Sri Lankan educational institutions utilize manual, paper-based systems to manage information which are more time and money consuming. It also reduces the accuracy and work efficiency. Nowadays the commercial world is fast reacting to the growth and potential in data science and as a result, data mining is getting much attention from many researches at present, and data mining assists to discover patterns within enormous amounts of data, stored in databases and data warehouses. Therefore, adapting these techniques will help to find interesting patterns to predict the student performance and to find the grades of students based on their examination results. Through this review paper, an effort is made to investigate a best data mining technique to quantify the student performance to provide benefits for academic staff, administration staff and students. The prediction on performance will provide more precise results and students may receive more accurate predictions which may help to make important decisions in their careers. Most importantly, this will reduce the workload of the administration and will surmount many challenges pertaining to the scholastic field providing a user-friendly environment.

Keywords- Data Mining, EDM, Classification

I. INTRODUCTION

Education system forms the backbone of every nation. Hence, it is important to manage education related processes to provide a strong educational foundation securing the future for everyone. Today educational institutions are not limited to imparting education alone, but also adapting the latest trends in IT to manage and serve the institution resources efficiently to improve the quality of education. At present, numerous studies are taking place in data-mining field. Educational Data Mining is one of the main research fields and also known as EDM. It aims at developing and using algorithms to improve educational results and explain educational strategies for further decision -making. This paper discusses some of the data mining algorithms applied on educational related areas. These algorithms are applied to extract knowledge from educational data and study the attributes that can contribute to maximize the performance.

One of the biggest challenges faced by educational institutions are the exponential growth of educational data and how to apply this data to improve the quality of managerial decisions. Educational Institutions would like to know, for instance, which students will enrol in particular course programs, and which students will need assistance for graduation. Through the analysis and presentation of data collected, on data mining process will help effectively to address the challenges of these students.

Data mining enables organizations to explore and understand hidden patterns in a vast range of databases by using their current reporting capabilities. And these uncovered patterns are then incorporated into data mining models and applied to predict individual behaviour and performance with high precision. In this way, resources

and staff can be allocated by institutions more effectively. Data mining may also, be able to efficiently allocate resources with an accurate estimate of how many students will take necessary actions before he or she drops out.

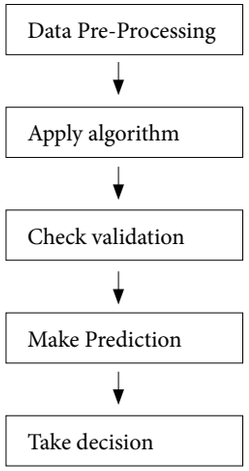


Figure1. Educational data mining Process

Generally, Educational data mining process mainly consists of four stages. During the first phase of the process, the relationships in between data is found by using a training data set. This is done by applying several data mining algorithms such as clustering, classification and association rule mining. In the second phase whatever the relationships identified in the data pre-processing step is checked for validation. As the third step, the validated relationships are used to make predictions. Finally, decisions are taken on the prediction results to filter data.

EDM process facilitates many users, specially administrators, educators and researchers. All the stakeholders will be benefited by data mining as it will ease their work in decision - making, do changes in teaching processes, allocating resources for institutions. So, the educational processes in institutions may provide best results.

II. LITERATURE REVIEW

Educational field has realized the importance of using data mining especially when examining students 'learning performances. Data mining can be defined as discovery of knowledge, which involves searching for unexpected but interesting patterns within large amounts of data, normally stored in databases and data warehouses.

Recently, data mining focused on analysing educational related data to develop models to improve student's learning experiences and enhance institutional efficiency. Therefore, data mining help educational institutions to provide high-quality education for their students. Applying data mining in education also known as educational data mining (EDM), which enables to better understand how students are learning, and determine how it improves educational outcomes. (Maqsood Ali,2013)

Student related information management, in an educational institute becomes tiresome each year, so it's better to have an automated system for managing such information. (Dalgade, et al., 2016)

Authors (NWOKE and IGBOJI, 2015) have developed a powerful offline and online school management software which plays vital roles in gaining competitive advantages. In addition, they explained that with proper planning and management, student records can be a valuable resource for many people, ranging from parents and local school officials to researchers and policymakers.

Educational data mining is emerging as a research area with a collection of computational and psychological methods and research approaches to understand how students learn. Therefore, if the academic planners can organize a conference for final year students then it will improve the overall productivity. (Dwivedi and Singh, 2016) provide an overview of the survey of student performance which helps academic planners in an institute to guide students to choose their right carrier in which their talents and commands are good. They have suggested that it's better to prefer algorithm like decision tree induction, association, logistic regression and naïve Bayes to filter educational data for placement predictions. (Dwivedi and Singh,2016)

(Bhardwaj and Pa, 2011) conducted a notable research on data mining using naive Bayes classification algorithm to predict the performance of students with 13 classification variables. According to his research observation, Data mining techniques are applied for vast amount of data to discover hidden patterns and relationships helpful in decision - making. Bayesian classification is one of the most useful method of data mining used on student databases to predict the students' performance on the previous year database. It helps students and the teachers to improve the class of the student and to identify students who need special attention to reduce the failing rate and to take appropriate actions at the right time. Moreover, that study shows university students' performance do not always depend on their own effort but, also depend on social, psychological and other environmental factors.

Evaluation is an important element in teaching and learning. The arrival of the Internet and related technologies have made online assessment systems feasible and popular in education and training. Evaluations can be formative or summarized. Computer based interactive multiple-choice question exams are proposed for formative evaluations while essay type exams conducted for summarize assessments. (Gogri et al., n. d) proposes a system of performance assessments for students. And the data generated from these evaluations are used for data mining. Classification technology C4.5 is used in decision tree showed that formative assessment, leading to better development of students while summarize assessment compels students to focus on how many marks have they secured rather understanding the content they are studying. Therefore, at the end they are suggesting that formative assessment strategies are more effective in improving quality of students learning and leading to better development of students.

There are various open source tools which are specialized for data mining. Some of them are Weka, RapidMiner, Orange, Knime, DataMelt, etc.(Kabakchieva, 2013) carried a research project using the WEKA software which is based on the C4.5 decision tree algorithm. As it is freely available to the public and is widely used for research purposes in data mining field. This research project reveals the high potential of data mining applications for university management by analyzing the performance of different data mining algorithms. And the results achieved from the research revealed that the decision tree classifier (J48) performs best followed by the rule learner (JRip) and the kNN classifier. The Bayes classifiers are less accurate than the others.

(Nithya et al., n.d.) has done a survey on educational data mining. They have discussed every algorithm which used in education mining. And have found that these algorithms show a remarkable improvement in strategies like course outline formation, teacher student understanding and high output and turn out a ratio.

(Lan and Li, n.d.) in 2011 published a research paper to improve the association rule mining algorithm. This algorithm mines the association rules of the courses, identifies the relationship of students, selected courses and provide basis for planning and curriculum classification. At the end of the experiment they have concluded that the association rule mining algorithm not only achieve the function of data mining, but also increases the mining speed to a certain extent and its simple and easy to implement.

(Baepler and Murdoch, 2010)stated that the new directions of academic analytics and data mining will produce new opportunities for collecting, analysing and reporting student's data. And those data can be used to redesign the course content, assessments etc. After series of experiments they have concluded that data mining is truly effective, influence curricular advancement and provide instructional choices for both students and the faculty.

(Yu et al., 2010) shows that data mining techniques can be used to study the factors that influence the retention of a university student. They have explored this using three data mining techniques namely, classification trees, multivariate adaptive regression splines(MARS) and neural networks. The data mining tools used in this study are used to provide some vision into various aspects of student retention that were not revealed before and encouraged researchers to investigate further.

(S. Abu-Oda and M. El-Halees, 2015) applied different data mining techniques with the purpose of examining and predicting students' dropouts through the university programs. They have used Decision Tree(DT) and Naïve Bayes(NB) classifiers and found out that accuracy of NB is higher than DT. This study also discovered hidden relationships between student dropout status and enrollment persistence by mining some frequent cases using F-P growth algorithm. Finally, they have concluded that mastering algorithm analysis courses a great effect on predicting student persistence in the major and decrease student likelihood of dropout.

(Agarwal et al., 2012) has taken student data from a community college and different classification approaches have been performed and a comparative analysis has been done. They have established that Support Vector Machines(SVM) as the best classifier and Radial Basis Kernel as the best choice for SVM. They had analyzed the

data available on a student's academic record and student likelihood in terms of placement may be predicted based on the admission test results, quantitative ability marks and verbal ability marks by using decision tree approach and adopting decision rule approach. And finally concluded that Data Mining could be used to improve the process of business intelligence including the education system in order to enhance the efficacy and overall efficiency by optimally utilizing the resources available.

III. METHODOLOGY

Education in any institution can be upheld by applying data mining technology, because Data mining helps to improve the standards and efficiency of the educational systems.

Various algorithms and techniques are used for knowledge discovery from databases. Classification, Clustering, Association Rule Mining and Sequence Analysis are some of the models used to implement a model.

Hence, identifying research questions will help to identify the scope of the study and to carry out a systematic review to support the researches.

A. Research Questions

Research questions are important to understand the study of predicting student's performance.

Two most important research questions are suggested in this study are:

- Q1: What are the important parameters, or attributes that are used to predict the student's performance?
- Q2: What type of algorithms are used to predict student's performance?

Next section of the study will discuss the proposed research questions that will be useful to predict students' performance.

B. The important parameters used in Student's Performance Prediction

When predicting student performance, the main predicated parameter will be their examination marks or the average score.

Before applying an algorithm, data has pre-processing step. If we assume, score variable have five distinct categories as grade" A"," B","C"," S" and" F". Students those who have scored in the range of 85 and 100 are belong to grade A, scores range between 75 to 84 belong to grade B, grade C in the range between 65 to 74, while S in the range between 55 to 64 and F in the range below 55.

To build a model that would classify the students into the five classes depends on the data collected and different algorithms are applied for predictions such as decision trees, Bayesian networks etc. We can train different models to predict student performances. In this way we can test the accuracy of each model.

C The prediction algorithms used for student performance

1) Naive Bayes Classifier

Bayes Classification is used to estimate the probability of a certain property in the data set. Naïve Bayes Algorithm is descriptive and a predictive type of algorithm. And it is easy to use from other approaches as, it only requires a small amount of training data to evaluation, because only one scan of the training data is required.

The Naïve Bayes algorithm classifies the cases by considering the independent effect of each attribute to the classification. And the ultimate precision is determined by the results achieved.

2) Decision Tree Classifier

Decision tree is a tree like structure that is used to describe the relationship between properties and their significance.

Decision tree algorithm is easy to understand and interpret. And works well for both quantitative and complex categorical data.C4.5 can be used with decision tree algorithm to obtain information which requires more attention.

3) Rule Learners Classifier

There are two algorithms, OneR and JRip. OneR generates one-level decision tree which is a simple and cheap method with high precision for characterizing the structure in data. And JRip implements the RIPPER algorithm where classes are studied in increasing size and generate set of rules for the class using incremental reduced-error pruning.

D. Logistic Regression

Logistic regression is used to describe data and to predict the dependent variables that explains the relationship between one dependant binary variable and one or more nominal variables ordinal, interval or ratio-level independent variables.

Table 1. Performance comparison between the classifiers

Author	Technique	Result
Oktarani Nuruk Pratiwi 2013	OneR J48 Naïve Bayes	78.66% 79.61% 76.75%
Ajay Shiv Sharma,S.S 2014	Logistic Regression	83.33%
Vikas Chirumamilla, B.S 2014	C4.5	77.78%

IV. DISCUSSION

In the field of education, for predicting the student performance, we can use data mining technology. It helps to make decisions on exam results and many more related to students. And lecturers can guide the students to the correct path to improve their career after analysing results.

From the review carried on educational data mining showed that the accuracy of Logistic Regression is higher

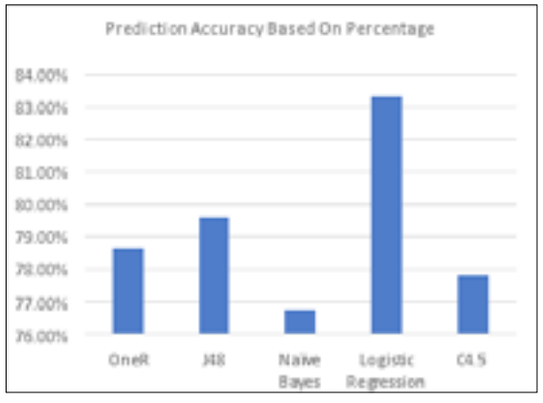


Figure 2. Distribution of Algorithms: Performance of the algorithms

than all the other algorithms. Therefore, it's better to use Regression and decision tree classifier as it performs best; followed by the rule learner (JRip) classifier to predict the student performance.

V. CONCLUSION

This paper analyses different data mining techniques to predict the student performance and find the grade of students. When predicting student performance, the examination marks or the average score have been used as a main data set by many researches. According to the research review, accuracy of logistic regression is higher than all the other algorithms for predicting student performance. Therefore, it's better to use logistic regression to predict the student performance. Further, this study will help students to make important decisions in their career path and lecturers will be able to identify weak students who need special attention to reduce the failing rate. Therefore, if all the scholastic centres can use data mining techniques in their resource management systems, it will improve the overall productivity and will enhance the future of students, improving learning and teaching processes. In Conclusion, I hope this review paper inspires the researchers to explore further on educational data mining and help to measure the student performance in a systematic way.

REFERENCES

Agarwal, S., Pandey, G.N., Tiwari, M.D., 2012. Data mining in education: data classification and decision tree approach. Int. J. E-Educ. E-Bus. E-Manag. E-Learn. 2, 140.

Baepler, P., Murdoch, C., 2010. Academic Analytics and Data Mining in Higher Education. Int. J. Scholarsh. Teach. Learn. 4. https://doi.org/10.20429/ijsotl.2010.040217

Bhardwaj, B.K., Pa, S., 2011. Data Mining: A prediction for performance improvement using classification. IJCSIS Int. J. Comput. Sci. Inf. Se Curity 9.

Dalgade, P.D.M., Panday, A., Negi, G., Popat, N., 2016. A RESEARCH PAPER ON STUDENT INFORMATION AND SCORE MANAGEMENT SYSTEM (SISMS) [WWW Document]. URL http://troindia.in/journal/IJACCCS/vol2iss2/13-16.pdf (accessed 10.28.17).

Dwivedi, T., Singh, D., 2016. Analyzing Educational Data through EDM Process: A Survey. Int. J. Comput. Appl. 136, 0975–8887.

Gogri, M.H., Shaikh, S.A., Iyengar, V.V., n.d. Evaluation of Students Performance based on Formative Assessment using Data Mining - pxc3886623.pdf [WWW Document]. URL http://research.ijcaonline.org/volume67/number2/pxc3886623.pdf (accessed 9.21.17).

Kabakchieva, D., 2013. Predicting Student Performance by Using Data Mining Methods for Classification. Cybern. Inf. Technol. 13. https://doi.org/10.2478/cait-2013-0006

Lan, A., Li, J., n.d. College information system research based on data mining [WWW Document]. URL http://www.icmlc.org/icmlc2009/080_icmlc2009.pdf (accessed 10.28.17).

Nithya, D.P., Umamaheswari, B., Umadevi, A., n.d. A Survey on Educational Data Mining in Field of Education.

NWOKE, B.O., IGBOJI, K.O., 2015. Automated School Management System – Recipe for Viable Educational System in Developing Countries. Int. J. Eng. Trends Technol. IJETT 25.

S. Abu-Oda, G., M. El-Halees, A., 2015. Data Mining in Higher Education : University Student Dropout Case Study. Int. J. Data Min. Knowl. Manag. Process 5, 15–27. https://doi.org/10.5121/ijdkp.2015.5102

Yu, C.H., DiGangi, S., Jannasch-Pennell, A., Kaprolet, C., 2010. A data mining approach for identifying predictors of student retention from sophomore to junior year. J. Data Sci. 8, 307–325.

DATA SECURITY SYSTEM FOR CHAT APPLICATIONS USING CRYPTOGRAPHY, STEGANOGRAPHY AND IMAGE PROCESSING

SC Mataraarachchi[1] and N Wedasinghe
Faculty of Computing General Sir John Kotelawla Defence University, Rathmalana1, Sri Lanka.
[1]*shashikalac38@gmail.com*

Abstract- The privacy plays a major role in the personal life. Due to the vast development in communication technology, people have the privilege to perform various operations seamlessly in their day to day life. But at the same time, privacy of some of those things such as confidentiality of communication is lacking due to the actions of some other parties. A message sent by the sender(s) should only be revealed by the intended recipient(s). Due to privacy issues, from the past, people have used various secret methods to preserve confidentiality of their information. Out of those, one of the developing science was steganography. Steganography is a method of encryption that hides data among the bits of a cover file, such as a graphic or an audio file. To read an encrypted file, one must have access to a secret key or password that enable them to decrypt it. At present this is widely used in various areas to secure valuable information. But in most existing systems, their security has not been trusted. Under this research, proposed a secure, flexible steganography mechanism to encrypt highly confidential messages that avoid them from being accessed by an unauthorized party.

Keywords- Data Security, Steganography, Encryption, Communication Technology, Confidentiality

I. INTRODUCTION

Instant messaging is a real-time communication medium that has grown increasingly popular for both social and professional use. In the Military sector, messaging can be used advantageously in scenarios where phone use is not possible or appropriate, for an example when communicating with geographical distributed teams, or for technical discussions in which the sending of URLs or operating system commands can be instantaneous. However, the military forces would be more concerned about security issues than a casual user who uses the software to keep in touch with friends because very critical & confidential information is passed among military officers.

One major disadvantage of most instant chat applications is that they are prone to security attacks. For instance, Yahoo messenger is vulnerable to security attacks when instant messages are sent between a source and a destination machine. The reason is the fact that, messages which are sent over the network is in a plaintext format with no encryption and decryption protection, thereby enabling intruders with no privileges to gain authentication sequence to alter the message content and make modifications to the message stream, learn the traffic pattern and cause denial of service.

The idea behind this research is to implement a simple messaging application that users can use to communicate sensitive information, which is highly secured with advanced stenographic mechanisms and encrypted on the front end to disallow outsiders from extracting secret information and at the same time succeeds in being unsuspicious and easy to use.

This system can be divided into 3 main parts. They are the front end, intermediate services and the back end. The front end consists of the interface and functionalities

related to the actions performed by the user. The intermediate services are the services provided by the server. Finally the back end consists of the services provided to the other user. The intermediate services allows to manage the database activities and encryption and decryption. The system will be using cloud server. At the other end there will be another interface for the recipient to receive the message.

The system has to cover different aspects such as register and log users, generate a virtual keyboard of user own language, cryptographic encryption, generate bit map images, send and receive secure message, send images and location tracking. Objective of the system is to develop a reliable, secure and a highly accessible messaging application which uses advanced data encryption standards and image processing mechanisms in such a way that it would benefit military and government personnel on secret surveillance or military intelligence missions.

II. LITERARURE REVIEW

A. MSN Messenger

Windows Live Messenger is a deprecated instant messaging client developed by Microsoft for Windows, Java ME, and S60 on Symbian operating system. It connected to the Microsoft Messenger service. Windows Live Messenger uses the Microsoft Notification Protocol (MSNP) over TCP (and optionally over HTTP to deal with proxies) to connect to Microsoft Messenger service. The American online service developed a buffer overflow bug, which causes it to execute a bit of machine code sent by the server. When this code runs, it determines if the client is MSN ID and sends a message back to verify the client.

The disadvantage of this application is the software has only allowed connections to its own service, requiring a Windows Live ID.

B. Yahoo Messenger

Yahoo Messenger is an advertisement-supported instant messaging client and associated protocol provided by Yahoo. It allowed Yahoo Users to create public chat rooms, send private messages, and use emoticons. On October 13, 2005, Yahoo and Microsoft announced plans to introduce interoperability between their two messengers, creating the second-largest real-time communications service worldwide. This allows Yahoo and Windows Live Messenger users to chat to each other without the need to create an account on the other service, provided both contacts use the latest versions of the clients. Yahoo messenger is designed to be compactable with windows, Mac OSX and Linux/Unix environment with their Operating system versions respectively.

Though it has some advance features, it was not possible to talk using the voice service among both messengers. According to thorough statistics, yahoo messenger has the weakest security features out of the two major IM providers (MSN and AOL), as it does not encrypt usernames and passwords, thereby running risk of data interception when the user logs onto the system.

C. Facebook Messenger

This is sometimes abbreviated as "Messenger". This app is an instant messaging service and software application. Over the years, Facebook has released new apps on a variety of different operating systems, launched a dedicated website interface, and separated the messaging functionality from the main Facebook app, requiring users to use the web interface or download one of the standalone apps. Users can send messages and exchange photos, videos, stickers, audio, and files, as well as react to other users' messages and interact with bots. The service also supports voice and video calling. The standalone apps support using multiple accounts, conversations with optional end-to-end encryption.

It uses data as its fuel and will not work when offline, like MMS and Text Messaging system. It can be taken as a big disadvantage of this application.

D. WhatsApp

WhatsApp is a freeware, end-to-end encrypted cross-platform instant messaging and Voice over IP (VoIP) service. The application allows the sending of text messages and voice calls, as well as video calls, images and other media, documents, and user location. The application runs from a mobile device though it is also accessible from desktop computers. The service uses standard cellular mobile.

The disadvantages of this application is, while changing to a new device using the same number, the existing chat stored in the old device cannot be retrieved into the new device incase of not having a chat backup.

E. Viber

Viber is an instant messenger app and communication tool for mobile devices. It is one of the many apps out there and is quite far behind the major players in the market such as WhatsApp and IMO. In order to remain in the game, it gambles on its free high-quality video and voice calls. It is a good app for video calls, with decent quality given all necessary conditions for good VoIP calling are there and scores quite high on Google Play and Apple App Store.

Though it is a good app for chatting, some unwanted messages like images and videos may take all the space in the device and also storing the chat backup mat be messy and takes time.

III. PROPOSED SYSTEM SOLUTION

There are various methods for data hiding like the Spatial Domain, Frequency Domain and Compressed Data Domain. Among them, this system has used the spatial domain. In this method, the image pixels in the spatial domain are arranged in order to incorporate the data to be embedded. This technique is simple to implement. It offers a high hiding capacity. The quality of the image in which the data embedding is done can be easily controlled.

This system uses AES encryption algorithm as the solution. A cipher in AES has a variable block length and key length. AES comprises of three block ciphers AES-128, AES-192 and AES-256. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128, 192 and 256 bits respectively.

According to the Symmetric Cipher Model of the system, the encryption algorithm performs various substitutions and transformations on the plaintext. The secret key is also input to the encryption algorithm. The key is a value independent of the plaintext and of the algorithm. The algorithm will produce a different output depending on the specific key being used at the time. The exact substitutions and transformations performed by the algorithm depends on the key. Cipher Text is the scrambled message produced as output. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different cipher texts. The cipher text is an apparently random stream of data and, as it stands, is unintelligible. Decryption Algorithm is essentially the encryption algorithm run in reverse. It takes the cipher text and the secret key and produces the original plaintext.

IV. METHODOGOLOGY

The system will be using an encryption algorithm to transform the user message (plain text) into encrypted text (cypher text) and it will be using a special formula for encryption purposes. Then mapping these cipher text characters with pixels with different RGB values so that it would form a random image. This image is the element that will be sent to the recipient. Even if an intruder retrieves the conversations, it will only be able to see a meaningless set of images. So in this way the system could effectively disguise the message preserving the confidentiality. When the message reaches the destination it will be retransformed into the encryption text and the cipher text will be decrypted using a decryption algorithm to provide the recipient with the original message.

After the application is setup on the android platform the user can access the login page for authentication. Then the relevant user name and password should be entered in order to login if the user already has an account. If not user can create a new account and afterwards login. After authentication is successful the user will be taken to the home page where the user will be able to see the other users who are currently online. Then the user can click another person's account who he/she wishes to chat with.

This enables a secure connection between the two users. Then the user will be taken to the messaging page, and there will be an option provided to the user as to alter the key mapping by assigning preferred keys by replacing default characters on keyboard keys. And when the user has finished assigning the keys he can click the generate button. Then the key map will be sent to the other user. Afterwards user can type the text and encrypt it and send it to the other user. In the user end the encrypted text will be decrypted and shown to that user after resolving the relevant mapping.

V. TECHNOLOGY

This system is mainly based on Java and Android. By using java, it will reduce extra cost that need to be bared for purchasing some components and tools. Because java is open source and all components are free to use. For development of this system java with NetBeans platform will be more suitable. Java is fully object oriented by design and more flexible in handling. Java is used in programming with NetBeans platform which provides an easier and flexible environment for the programming purposes and GUI developments. Java is platform independent, so it will be able to reuse the system components in effective manner at a later time. If the platform is going be changed then, no redevelopment will be required for the system. So, it will be more beneficial for the users if they are intending to change the system requirements in the future.

VI. SYSTEM DESIGN

This system can be divided into 3 main parts. They are the front end, intermediate services and the back end. The front end consists of the interface and functionalities related to the actions performed by the user. The intermediate services are the services provided by the server. Finally the back end consists of the services provided to the other user. The intermediate services allows to manage the database activities and encryption and decryption. The system will be using cloud server. At the other end there will be another interface for the recipient to receive the message.

The system has to cover different aspects such as register and log users, generate a virtual keyboard of user own language, cryptographic encryption, generate bit map images, send and receive secure message, send images and location tracking. Objective of the system is to develop a reliable, secure and a highly accessible messaging application which uses advanced data encryption standards and image processing mechanisms in such a way that it would benefit military and government personnel on secret surveillance or military intelligence missions.

This is intended to develop as an instant chat application known as Secured Java Chat Messenger (SJCM). SJCM is an instant chat application that provides an intuitive and reliable way of exchanging instant messages over a network using two or more computers. However SJCM (Secured Java chat messenger) provides an assurance of network security where plaintext are transformed into cipher text (Unintelligible message), thereby making it difficult for a cryptanalysis or an intruder attempts to alter message content, to make modifications to module (independent) and to the client chat application module (dependent). The server chat application module provides a graphical user interface with settings and options which enables an effective and secured exchange of instant messages between two or more communicating entities. Similarly, the client chat application module also contains a graphical user interface and provides a reliable and secured communication.

Secured java chat messenger would be built upon the ideology of the client server architecture model. However, these two modules (server and client chat module) can communicates with each other on a network upon a connection establishment which would be dependent on the host computer's port numbers and IP addresses. Each model would contain an encryption and decryption scheme using the encrypt the plaintext into a cipher text with a private symmetric key before making the transmission on the network and decrypt the cipher text into a plaintext with the same private symmetric key at the receiving end of the communication.

A symmetric private key is a unique pass code that is used to secure an encrypted plaintext at the transmission end of the communication and used to decrypt the cipher text into a plaintext at the receiving end of the communication. The key will be generated by the user according to his/her preference when they are assigning keys to the keyboard. Each key assigned to be relevant for the keyboard click will be converted into Unicode values and they will be the RGB values relevant for a single pixel. After developing all this pixels on a plane, can send it to the other user. Then that user can compare parts of the message with the received key map in order to determine the original message.

VII. SYSTEM FEATURES

Receiver can view the plain text of the message with a click of the button provided with chat message. They can click the button again to display the text in secret language.

Figure 1. Process flow of the system

Specialized encryption and decryption techniques are included in the application and the users have to login with the password they provided for registration, every time they use the application.

Auto destruction of viewed messages after a user defined time period. A chat can be started only with a person who has accepted the chat request. Session time-out will occur when kept idle for some time and re-login will be required to pursue.

The users can delete messages from application, from server and from receiver's chat history. Messages can be sent with a time limit so that it self-destructs after time-out.

Reducing database access time by using SQLite for some frequent data needs. Navigation on system would be user-friendly and accessible. System has simple and interactive user-interface. Reduced data usage on the application.

System should be available at all times and the central database should be updated each time user data interaction happens on the system.

VIII. SYSTEM ARCHITECTHRE

The architectural design of the system is elaborated in this section. The key components are illustrated further more. Overall system will be split in to three layers named presentation layer, application layer and data link layer. Based on this structure, the components will be divided in to separate modules to manage the operations of all components.

The presentation layer shows the interactions with users by controlling interfaces to display requested information and accept the inputs provided by the end user. Information gathered by presentation layer will be provided to the application layer in order to manipulate according to the given instructions.

Application layer can be named as the heart of the overall system and whole encryption logics and processes of the system will be executed at this layer in order to gain the proposed objectives of the system. This layer will interact between application layer where the interfaces are operated and data layer where the information is stored. Data gathered by user inputs or by other processes will be executed according to the predefined operational instructions at this layer. Data layer controls the data storage operations of the overall system where the database management applications are running. Also system uses log files to keep messages.

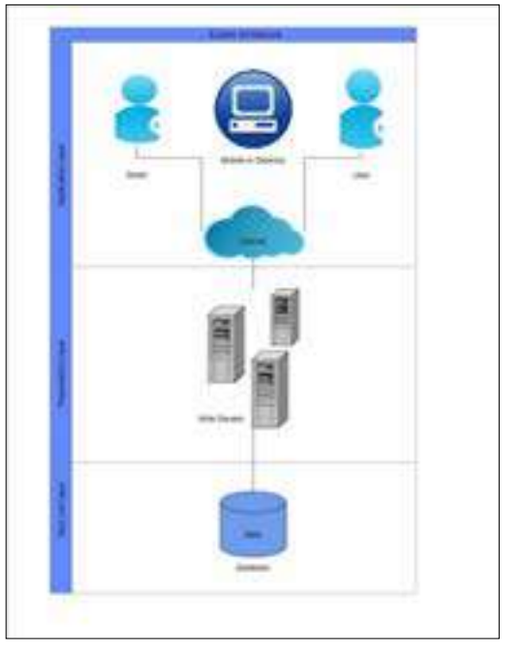


Figure 2. Three layer architecture of the system

IX. CONCLUSION

The proposed system is a low-cost, secure, flexible steganography mechanism for all users to encrypt highly confidential messages that avoid them from being accessed by an unauthorized party. This system has revolutionized the instant messaging service, minimizing the drawbacks of the existing systems and improving the performances of the technological features. With the help of the above technological features, allows the system to recognize multiple panoramas in unordered image sets, and stitch them fully automatically without user input. It has put forth a new system which combines text cryptography and image steganography which could be proven a highly secured method for data transactions in the near future.

ACKNOWLEDGEMENT

I would like to thank my supervisor Ms. N Wedasinghe for her constant support & guidance. Her active cooperation & involvement have helped me through the various stages of project development.

REFERENCES

Madden, M. (2003) America's online pursuits: the changing picture of who's online and what they do. Washington D.C.: Pew Internet & American Life Project.

Adams, C. "Simple and Effective Key Scheduling for Symmetric Ciphers." Proceedings, Workshop in Selected Areas of Cryptography, SAC' 94. 1994.

Bellare, M.; Kilian, J.; and Rogaway, P. "The Security of the Cipher Block Chaining Message Authentication Code." Journal of Computer and System Sciences, December 2000.

Beth, T.; Frisch, M.; and Simmons, G. eds. Public-Key Cryptography: State of the Art and Future Directions. New York: Springer-Verlag, 1991

Biham, E., and Shamir, A. Differential Cryptanalysis of the Data Encryption Standard. New York: Springer-Verlag, 1993 Daemen, J., and Rijmen, V. "Rijndael: The Advanced Encryption Standard." Dr. Dobb's Journal, March 2001

ElGamal, T. "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms." IEEE Transactions on Information Theory, July 1985

Jueneman, R.; Matyas, S.; and Meyer, C. "Message Authentication." IEEE Communications Magazine, September 1988.

Kohnfelder, L. towards a Practical Public-Key Cryptosystem. Bachelor's Thesis, M.I.T., May 1978 Rapp, D. (2002) I've got to get a message to you. Technology Review, 105, 8, 88.

Broughton, K. (2002) Usage and user analysis of a real-time digital reference service. The Reference Librarian, 79/80, 183-200.

"Automatic Panoramic Image Stitching using Invariant Features", Matthew Brown and David G. Lowe of Computer Science,University of British Columbia,Vancouver, Canada.

Jones, Chris (August 17, 2010). "Windows Live Essentials 2011 beta refresh". Microsoft. Retrieved August 17, 2010

NavitaAgarwal, Himanshu Sharma "An Efficient Pixel-shuffling Based Approach to Simultaneously Perform Image Compression, Encryption and Steganography", IJCSMC, vol. 2 issue 5, pp. 376385, May 2013.

Mohammadi S., Abbasimehr H., "A high level security mechanism for internet polls", ICSPS, vol. 3, pp. 101-105, IEEE, 2010.

Gary C.Kessler, "An Overview of Cryptography: Cryptographic", HLAN, ver. 1, 1999 2014.

MilindMathur, AyushKesarwani, "Comparison between DES, 3DES, RC2, RC6, BLOWFISH and AES", NCNHIT vol. 1 143-148, 2013.

Ashwin S., "Novel and secure encoding and hiding techniques using image steganography: A survey", ICETEEEM, vol. 1, pp. 171-177, IEEE, 2012

Chanu Y. J, "A short survey on image steganography and steganalysis techniques", NCETAS, vol. 1, pp. 52-55, IEEE, 2012.

Chung, H. M., & Silver, M. S. (1992). Rule-based Expert Systems and Linear Models: An Empirical Comparison of Learning-by-Examples Methods. Decision Sciences.

Cios, K., Pedrycz, W., &Swiniarski, R. (1998). Data Mining Methods for Knowledge Discovery. Norwell, MA: Kluwer Academic Publishers.

WORKERS' ALCOHOL DETECTION AND PREVENTION SYSTEM

KC Yakupitiya[1] and Wijendra Gunathilaka
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]*kavinduyaku@gmail.com*

Abstract- Factories are industries of vital importance, and they operate with two main resources, machinery and manpower. When dealing with Machines, workers should work carefully as an error could lead to injury or loss of life and trade. The biggest industrial machine which operates with high power are critical because a small mistake upon those may lead to huge losses. In this paper an automatic system to detect alcohol ingesting of factory labourers and the engine shut down with a warning scheme is described. Though this is a common problem in many factory areas no one seems to focus on it in order to prevent the problem that could occur when dealing with machines and lives right away. Humans are always doing silly things. But taking alcohol and doing some heavy stuff in their working areas could lead them to huge losses even more than people could ever imagine. So, in this research paper, how to take control of these things and to prevent workers from alcohol usage in their own working area has been summarized.

Keywords- Alcohol Detection and Prevention

I. INTRODUCTION

So, when talking about alcohol addiction most of the grown human are alcohol addicts. According to 2015, the Nationwide Drug Use and Use Study (NSDUH), 86.4% of people over the age of 18, transferred that they had a drink at a certain time in their lives. But in here it does not directly focus on the addiction of alcohol and their negatives to a human body. In here our main focus is how the alcohol usage and those impacts directly influenced to the factory workers' lives and their misuse of tasks. Not only that, another main focus is to help the company owners to reduce the damages which could occur because of their working staff. So they will be able to reduce alcohol usage and by doing that they can reduce the damage and even increase the work performances so they could increase their brand name highly. Down here in the (figure 1) are some recent finding of alcohol addiction in our day to day world. So, you guys can clearly understand how much effect it kept on doing to a human. So, people could never say no to alcohol usage in the working area between human.

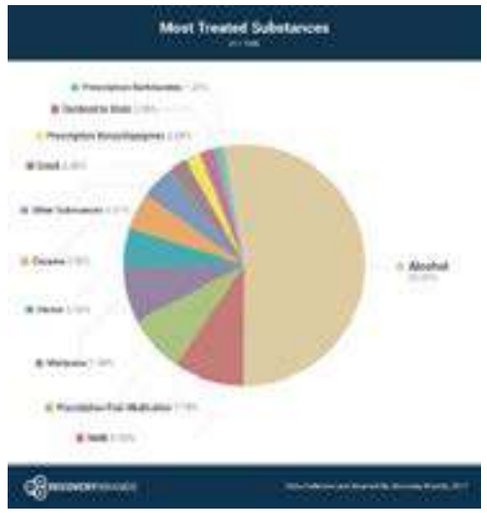


Figure 1. Usage of alcohol in worldwide prior to other things

As web applications are increasingly used to deliver though there are many systems and research papers has been

published regarding alcohol detection and prevention. 95% those are for the safety of the drivers who use alcohol. Their main goal is to prevent the accidents which occur because of ruthless driving while being drunk. Because so many systems were developed on driver's alcohol detection and prevention main focus was to touch a side which no one had ever touched earlier, in order to develop an alcohol detection and prevention system of workers.

Drinking of alcohol moves the psychological state of a person. Operating heavy machinery under the impact of alcohol, which can be led to self-injuries. So, in here, developers are going to set up the alcohol detection system at the front door where the workers enter. For the identified purpose, it is going to set it up right with the fingerprint or whatever the identification machine where workers use to mark their present to the working area. So, an MQ3 sensor is there in order to detect alcohol gas. Then uses a microcontroller-based circuit that contains of an alcohol sensor interfaced with it. Also, have a GSM modem and an LCD display. The whole scheme is powered by a 12V source. Alcohol sensor regularly runs in an uneven directions to check whether the worker is drunk. The alcohol direction sensor can intellect the level of alcohol and produce an output according to the alcohol that it felt.

The microcontroller interfaced to it reads the value and if it is found to be above the permissible range it goes into alerting mode. As soon as they enter the alarm mode, the microcontroller prevents the machine operated by the worker and displays the status of the alcohol alarm on the LCD screen. While on the exact same time, there will be a mobile alerting system for the manager of that area saying that this worker is drunk at this moment.

Now it was avoiding the disasters, the system now must submit a report to the applicable offices/owner of the industry, reporting on the incident sends. It will then inevitably send a text message via the GSM modem's official number to inform the situation so that the necessary actions can be taken from it. So basically, that is what will be going to get deployed by this system.

II. LITERATURE REVIEW

A. Early Developments.

The Alcohol prevention or detection systems has been dramatically increased and many research papers and review papers have been published. Some of them discussed the danger of the work environment, analyzing the security and the privacy risks of people.

Many research papers have been published all those years regarding alcohol detection and prevention system, but most of those are about the driver's alcohol detection and prevention state. Phani Sridhar et al (2014) has done a research based on Alcohol Recognition through Automatic Motorised Padlocking System: In Built (LDAMLS) has provided a specific concern about what alcohol can do to a man while driving etc. it's same like operating a machine. Operators got to stay focused on what they do. Here he clearly says what alcohol does to a human body and their mind. It could destroy human sensors somewhat than what people could ever imagine. So, he estate how dangerous it can be. Just like operating a machine, driving is also so similar to that task. Because both are machines and in order to take control of a machine first workers have to take control of their selves. But being drunk is a disaster, so worker could be in danger. That's why this alcohol detection and prevention system is going to developed for an office.

Not only that. M. K. Mishra et al (2013) has also done a research based on the Fuzzy Based Model for Accident Avoid System, which also a focus on controlling vehicle accidents in many ways. Here he has his focus on a different way, but towards the same target. Here he has industrialised fuzzy logic to make one paper, which operates the emanation of toxic gases, smokes inside the vehicle and dodge the drunken drive. Though the technique is dissimilar, it also uses sensors to notice smoke, air, and gas. Then the warning system go on to snooze to warning mode where many drivers could easily get their lives saved.

Mandalkar Rahul B et al (2015) has provided an experimental analysis of the same issue in some other ways and all are based on Liquor Detection and Accident Escaping Using Padlocking with Tracking. In here they clearly mentioned that lots of drunk drivers and accident were clearly not found by the police. So just imagine if the police can't catch drunk drivers how organizational owners can get a chance to catch their drunk workers who operating their machines. In that system, they proposed a system that would show the drunk drivers using alcohol sensor with the use of driver's breath, along with the driver's navigation. System also provide a GSM system that sends a message to the nearest police and provides

the appropriate data for the driver's alcohol detection application. As they use alcohol sensors, the driver tries to detect an irregular heart disease, which is why they are drunk drivers.

B. Early Findings of alcohol usage.

Conferring to a study done in the year 2016 around Canada and the USA (figure 2.0) exposes that nearly 50% of their occupied employees are addicted to liquor. So, you all can imagine the impact that can bring to their industry.

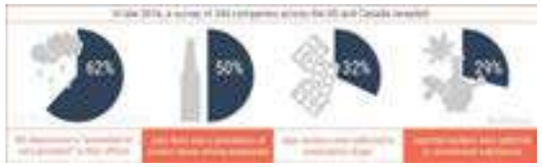


Figure 2. Survey was done in the USA and Canada about company workers alcohol addiction

Another survey was done by the Royal military in early 2016 and that state, even the most disciplinary places like Air force, Army, and Navy too addicted to the liquor so much (figure 3)

	Percentage
Royal Navy (inc. Royal Marines)	31.6
Army	19.9
Royal Air Force	24.6

Figure 3. (Royal Military based alcohol addiction while working)

Think of other office locations except these military places.

Not only that there is indication to show that the damage of skills begins with any important amount of liquor in the body.

For example, in a study on airline pilots who have to achieve routine tasks in a simulant in the case of 3 liquor tests; [5]

- Before the breakdown of any alcohol, 10% of them could not achieve all the processes properly;
- After attainment a blood alcohol attention of 100mg/dl, 89% could not achieve all the actions properly.
- Then 14 hours later, after all the liquor had left their schemes, 68% still could not achieve all the processes properly.

So, visualize the destruction that could bring to the organization by having a liquor in the body of their labours.

C. Existing Systems and Solutions.

Vaishnavi. M et al (2014) has introduced an alcohol detection and prevention system for drivers. It offers the most advance in the alcohol sector, a device that notices a change in the content of alcohol fuel in the air. The sensor will analyse the amount of alcohol vapours and give the user a sign of the amount of alcohol present. This device is more commonly referred to as an analyser; since it analyses the liquor content of a person's breathing. When someone used it, anyone can obviously see that it is a useful method.

Not only that, like in earlier mentioned researches Phani Sridhar et al (2014) has done a research based on Liquor Detection through Automatic Motor Locking System and it senses the air and smell so the system could capture alcohol detection and it will automatically lock the car and block it.

Kousikan1 et al (2014) has done a research based on Automatic drunken drive prevention system by which is integrated with the steering wheel. Ethanol has an advanced capacity to absorb infrared radiation. Therefore, the system use an IR sensor that is connected to send an address. Source IR LED-894 IR Power Head from Sensor Continuous. If you stop the IR beam from the fascinating alcohol light, the relay path will be triggered. This link shackle has changed the system of gas origin and fails the supply of engine fuel.

Just like earlier mentioned M. K. Mishra et al (2013) has also done a research based on the Fuzzy Based Model for Accident Prevent System uses smoke and smell. Although they all look different they all merge as against one goal. It is to detect the alcohol of the human body and avoid incidents that may arise. Just as the office worker's Alcohol Detection and prevention

III. METHODOLOGY

According to the above works of literature, it is easy to identify what are the pre-made systems for alcohol detection and prevention. Though every research seems to be about drunk driver's safety this is something different to all other paths. That is office worker's alcohol detection and prevention system. So, because of not having further details to carry on this research, there is a survey developed to collect more and more details. In that survey, there are 14 questions and sent it to people who have been working in the industrial area for some time.

So, in that survey, it has been declared as 75% male and 25% female. That is because lots of males are addicted to alcohol far than the females. But still the experiences of the females around the working area is much more important because they might have witnessed some of those incidents where their office mate get caught drunk while working. Then at the same time, there are some age limits like 16-25, 26-45 and 45 – more.

Then the next step is to check the number of people who take alcohol and who does not. But even someone who does not take alcohol must have some experience of it, while working because of their fellow mates who drink somehow. So, through this survey it was easy to collect the details, because the entire survey covered all those areas where workers drink or not, how much will it affect to the working area.

This survey covered all the areas like whether there is any alcohol detection system inside their office, if it is "what kind of system is it". Were there any incidents regarding alcohol usage in their office etc.

Likewise, in here there are many more important questions and it developed by using Google Form to collect the details, and it distributed between almost 50 people around the industry to capture their experiences and ideas. While at the same time there are sample two questions to collect their feedback regarding this system. So, they can give an idea of what this system should do and what are the things should be added etc.

So, the past week or so there were much more interesting ideas and feedback collected from various people and this is a simple method which used to collect information regarding this idea of developing alcohol detection and prevention system.

IV. RESULTS

When talking about results first there were inclusion of 75% males and 25% females in this survey. Then the survey added some age limits and the interesting thing is that lots of them are young teenagers between 16-25. As a percentage, it's 77.3% and other remaining 22.7% are more than 26 and above. When responding to these questions, it asked whether they are alcohol users or not, interestingly more than half of them are alcohol users. So, you can imagine the impact of alcohol towards human these days. As a percentage, it's 56.8%. So even from that point onwards, it says that this research is a worthy one.

Then the interesting part is that the survey added a question, saying have you ever seeing any of your mates being drunk while working and the response is 40.9% (figure 4.0) so that means lots of them are using alcohol while working without knowing what they do.

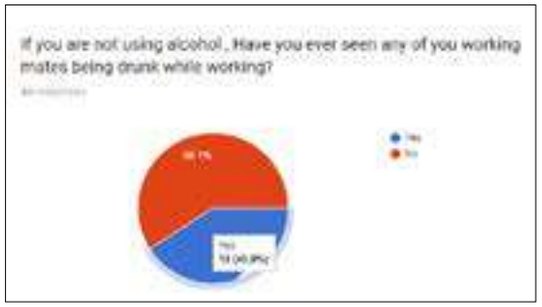


Figure 4. Regarding alcohol users inside the office

When asking whether they had witnessed any of their supervisors caught their drunk office workers in the office only 30.2% were saying yes and another 69.8% said no (Figure 5.0). So that means it's rare to get caught while working drunk inside the office.

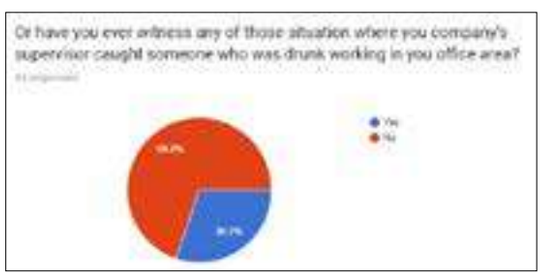


Figure 5. any incident of a supervisor attempts to capture their drunk workers

When it comes to the main questions whether there are any systems to detect alcohol users while working (figure 6), 79.5% of people said there is no exact system inside their office. So that motivated further to develop this system.

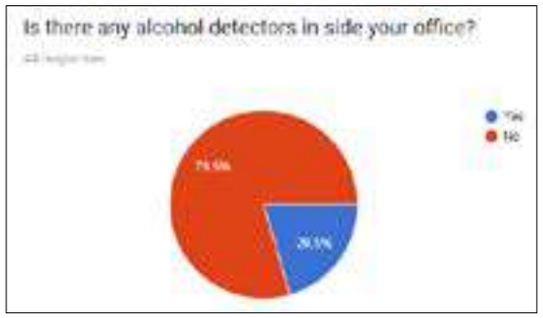


Figure 6. to check whether any alcohol detectors inside the office

When asking further they said that there is no way to check alcohol usage if they needed to. So, they said that there were only 59.6% times where their supervisors couldn't take actions against their drunk workers because they couldn't prove it. When asking whether it is a good idea to develop an alcohol detection and prevention system in their office almost 70% (figure 7) of people agreed and that indicates the need of alcohol detection and prevention system around an office areas.

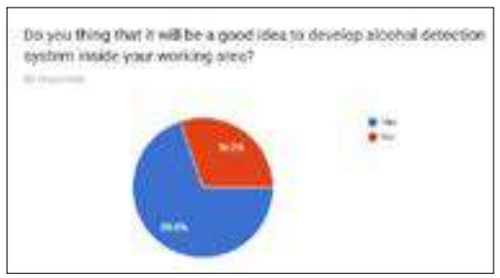


Figure7. To check whether it is a good idea to develop an alcohol detection and prevention system

When further asking about what were the problems which their companies faced regarding their drunk workers they said that they saw some workers fighting and inefficiency of their work performance, and even its hard to live around the office because of the smell and they are afraid that even they will face legal matters if there is any situation occurred etc. Then finally through the survey it gave them to provide feedbacks asking what the positives are and what are things should be further added in order to develop this system.

So, they said that the system has to be on alert for 24 hours and should keep track of their drunk employee's record. Then they state that the system must be updated with the technology. While not only with technology the mindset of the people should be corrected too. Some of them said it should be right there with the fingerprint in order to detect exact employee. So, in order to do that there will be a database to keep the track of their records according to the identification purpose.

- They can also say that it can bring safety, dangerous working location with damages and mortalities, especially in occupations where substantial machinery or motor vehicles are used.
- **Workroom affairs** with colleagues and clients and customers. Effects on liquor may affect performance, such as a person who is prejudiced by alcohol, may amateurishly treat customers and clients in working circumstances; Colleagues can be solid on other affected alcohol or "hangovers" The effects of alcohol are sheltered.
- **Low labour productivity** due to nonattendance in the short term and a lower value and quantity of work due to the poor choices making and the disruption of operations and co-workers to cover for labours affected by alcohol.
- **Workplace economy** with long-term alcohol-associated problem attributed to recompense and employer responsibilities; loss of services and employees; and the associated charges of replacement and training new employees can be happened and having alcohol detection and prevention system could stop those unwanted and uninterrupted things from occurring.

V. DISCUSSION AND CONCLUSION.

So, when taking all those methodologies, surveys and results final solution is to develop this alcohol detection and prevention system.

In order to do that there are some software and hardware specifications,

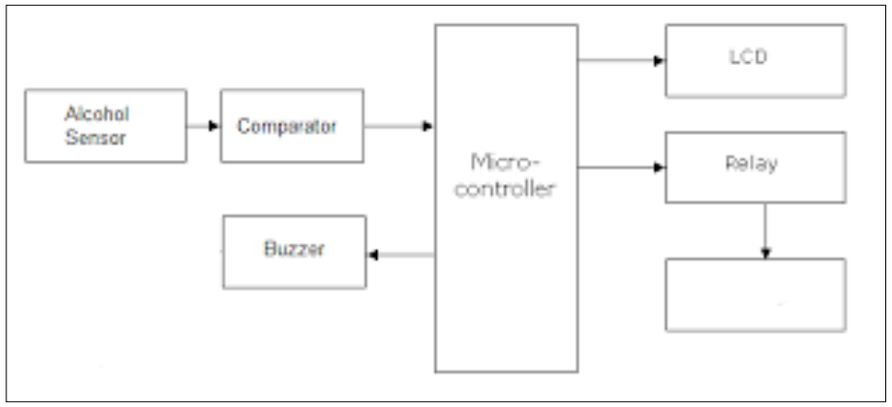


Figure 9. Small Chart How it Works

Hardware Specifications

- PIC Microcontroller
- Rectifier
- Regulator
- Transformer
- GSM Modem
- MQ3 Alcohol Sensor
- Buzzer
- LCD
- Motor
- Resistors
- Capacitors
- Diodes

Software Specifications

- MC Programming Language: C
- Android
- MYSQL

Therefore, the system use a microcontroller-based circuit that consists of an alcohol sensor (MQ3) connected to it. In addition, it have a GSM modem and an LCD display. The entire system is equipped with a 12V supply.

Then from feedbacks which is being received, the system must set it right up there with the fingerprint or any identification machine where workers enter into the working area. So the system could detect which person exactly is drunk at the exact time. So, in here the MQ3 Alcohol sensor will be set in front the entrance, so any person before entering the working area the detector will detect whether that person drunk or not. Through Alcohol sensor you can smell the alcohol level and the voltage according to perceived alcohol productions. The microcontroller interfaced to it reads the value and if it is found to be above the permissible range it goes into alerting mode. After entering the notification mode, the microcontroller prevents the machine operated by the worker and displays the status of the alcohol signal on the LCD screen. Also, it sounds a Bell to show the same. Now separate the car used to prove himself as a machine.

Now the Miss bite was avoided, the system now needs to submit a report to the appropriate authorities/industry owner to inform about the incident. It then automatically sends an SMS message to the GSM modem to an authorized number to inform about the situation so that the required action can be taken. And to talk about other workers there is a need of developing a database that will keep track of the worker's details so that supervisors can take steps accordingly.

VI. CONCLUSION

Alcohol detection and prevention is a must in the modern era because there are lots of accidents and problems which

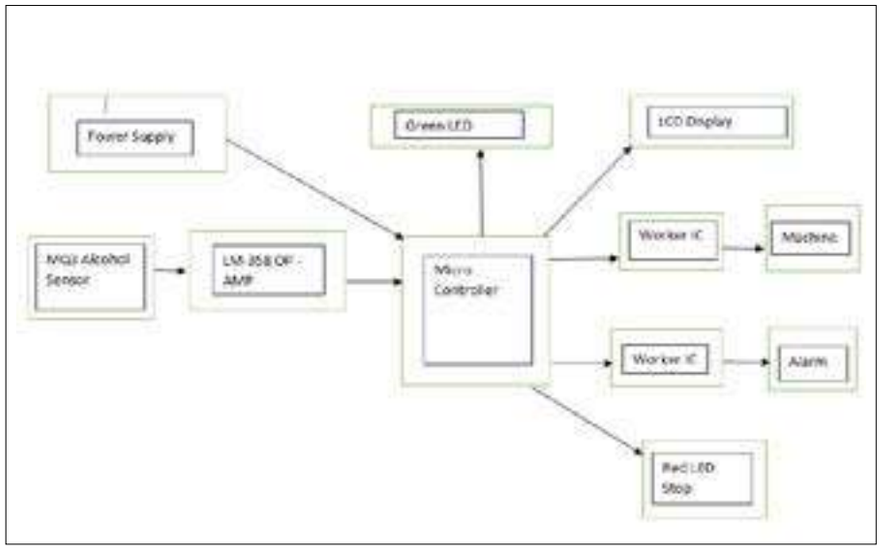


Figure 8. Small Chart How It Works

has been occurring in an office areas throughout the world. With the modern technology, developers can get many more ideas about how to develop an efficient alcohol detection and prevention system easily. Talking about the idea the arduino part has already been developed and it can be used for many vulnerabilities like helping police to detect drunk drivers, prevent alcohol usage in colleges, workplaces even when doing sports etc. Because this topic is a new topic to Sri Lanka and existing systems regarding alcohol systems are very hard to find.

ACKNOWLEDGMENT

The author would like to acknowledge the supervisor, Dr. (MRS.) Wijendra Gunathilaka for her massive support. Then (MR.) Ruvan Pathum, (MR.) Maithri Ranga Kulasekara and (Mr.) Nisal Hewagamage for guidance of path regarding all the hardware specifications. Finally the author would like to acknowledge all the supporting staff who taught and supported to complete this project.

REFERENCES

Sridhar. A, P., Susan. V, S., Chakravathi, and Teja. G, R. (2014). Liquor Detection through Automatic Motor Locking System: In Built (LDAMLS). [Online] Ijceronline.com. Available at: http://www.ijceronline.com/papers/Vol4_issue07/Version-2/E04702056061.pdf [Accessed 14 Apr. 2018].

M.K.Mishra, S.Mohanraj, T.Yazhini, K.Vijayasri, R.Gomathi Ijsrp.org. (2013). Fuzzy Based Model for Accident Prevent System. [Online] Available at: http://www.ijsrp.org/research-paper-0313/ijsrp-p15104.pdf [Accessed 12 Apr. 2018].

Rahul B1, M., Rahul N2, P., Manoj B3, S. and Valmik D4, G. (2015). Alcohol Detection and Accident Avoidance Using Locking With Tracking. [Online] Ijarcsms.com. Available at: http://www.ijarcsms.com/docs/paper/volume3/issue9/V3I9-0017.pdf [Accessed 7 Apr. 2018].

Vaishnavi.M , Umadevi.V , Vinothini.M , Bhaskar Rao .Y, Pavithra. S Anon, (2011). INTELLIGENT ALCOHOL DETECTION SYSTEM FOR CAR. [Online] Available at: https://www.ijser.org/researchpaper/INTELLIGENT-ALCOHOL-DETECTION-SYSTEM-FOR-CAR.pdf [Accessed 10 Apr. 2018].

Ias.org.uk. (2017) The effect of problem drinking in working areas Consumption - IAS. [Online] Available at: http://www.ias.org.uk/Alcohol-knowledge-centre/Consumption.aspx [Accessed 14 Apr. 2018].

M. KOUSIKAN1, M. SUNDARAJ2 GIAPJOURNALS.COM. (2014). AUTOMATIC DRUNKEN DRIVE PREVENTION SYSTEM. [ONLINE] AVAILABLE AT: HTTPS://GIAPJOURNALS.COM/INDEX.PHP/IJSRTM/ARTICLE/DOWNLOAD/188/182/ [ACCESSED 13 APR. 2018].

FAST AND ACCURACY PALM-PRINT RECOGNITION SYSTEM FOR LOW-QUALITY PATTERNS

EAIL Edirisinghe1 and RL Dangalla
Department of Computing and Information Systems, Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, Belihuloya
[1]isuru991@gmail.com

Abstract- Palm-print trait based biometric identification has emerged as a most powerful tool to recognize a person's identity. It is used in commercial and forensic applications. In common, it considers high 400 dpi (Dots per Inch) or more is high resolution and 150 dpi or less is low resolution. Earlier research projects have been showed that high-resolution palm images are capable to extract ridges, singular points and minutia points as features. Low-resolution images have capability to extract principal lines, wrinkles, and texture. Therefore, researches which are based on palm prints primarily focus on high-resolution palm images. Therefore the main purpose of this research was to provide a Fast and Accuracy Palm-print Recognition method for low-quality images using image processing with feature extraction. In this research, the database which was used for this is mainly consisted with the palm images which were collected from students. In feature extraction process, low pass filter was used to remove the noises of images. Elongated and tubular structures were enhanced in the noise removed images to highlight major lines by using Hessian-based multi-scale filtering. Segmentation process in the original image transforms the original image in to a binary image in which ridges area fully colored in one tone and the background in the opposite tone. Threshold binary image was applied to some morphological operations to extracting the better results. The palm-print matching process which is also called as the template matching was mainly based on the normalized cross correlation in Fourier domain (phase correlation). This process was done by pixel-by-pixel basis. Results of this process have shown a higher Genuine Acceptance Rate for lower False Acceptance Rate and False Reject Rate.

Keywords- Low Resolution, Principle Lines, Feature Extraction, Template Matching

I. INTRODUCTION

In the real world, in our day to day life, there are so many incidents which have to be faced to so many difficulties because of the lack of concentration due to the turbulent environment of people. As an example what would be happened when it is found that the door key has lost after coming home finishing the job. It could be annoying and finally, it can destroy your whole plan. In the present day, there is very important and valuable resource to find solutions for all of these problems. What would it be? It is progressive technology. With the technology, it is easy to solve these simple but important problems. So, when finding solutions to the above common problem, a simple thing which we always have can be used as the door key. It can be voice, hands, eyes, iris, and fingerprints or palm prints etc. These simple but important things can be used with the technology to solve problems as well as to increase the efficiency of our day to day life.

The using of biometric characteristics of a person to recognize him or her through the technology have been discussed over many years of time. But still it has not come very common and people do not see those technologies in real life in addition to fantastic films. The reason for this situation could be that people may not be able to bear the high expense for these technologies as the practical use of biometric methods are too expensive and there is a myth that it could not be done in real life and people consider as it is only a matter of future. Face recognition, iris scan, signature scan, keystroke scan, palm print are some popular bio-metric identifiers using in the present world. But these characteristics sometimes cannot be able to identify properly with dependence of low-resolution images, form of hair, facial expressions, sometimes it requires special devices and accurate positioning as well as a specialized training. Some technologies may harmful to the people's health. But this palm printing method has many unique advantages when comparing with other methods of biometrics.

- Age of the user is a minimum affected factor in palm printing method. But in face recognition method, it is a major factor to be considered.
- Palm prints contain more information than other methods and low-resolution devices can be used for this identification process of those information.
- It is not harmful to the people's health. Therefore many people prefer palm recognition despite iris recognition.

There are mainly two types of palm-print recognition system.

i. High resolution

ii. Low resolution

Human palm has unique distinguishable line patterns which can be used to identify people uniquely. This very large and internal surface of the human hand which is called palm, contains several specific characteristic features as principle lines, wrinkles, creases, and textures. Therefore palm prints are very robust to noise and unique to every individual.

Comparing with other physical characteristics, palm prints contains many useful and unique features. Those are as follows,

- Principle lines- These lines very little over time. Therefore both location and form of principal lines can be used for recognition of individuals.
- Wrinkles- Wrinkles form rich texture in the hand and, are thinner and more irregular in comparison to principal lines.
- Ridges- Ridges are spread all over to the hand and very thin. It needs high resolution to detect them.
- Delta points - Delta points are the center of the delta-like region of the palm print.
- Minutiae point- Minutiae points are the ridge characteristics, ridge ending, and ridge bifurcation.

A novel technique has been proposed by Zia-Uddin to extract principal lines, which is evident and stable features in palm-print images in two phases without using edge detection. In the preprocessing phase, normalization is done to enhance the contrast of the palm image. With the help of smoothing and median filters noise is removed. In last step of the phase, the palm image is converted to negative. Next in extraction phase Top-Hat filter and contrast enhancement are applied respectively on the negative image and noise is removed with connected component labelling after binarization. In this research, they have used IIT Palm-print Database version 1.0. [8]

Pawan Dubey has presented a new palm-print scheme which is called as Sequency code (SeqC). It exploits symmetric Gabor filtered responses at different orientations. Concatenation of the zero crossings of these symmetric and asymmetric Gabor responses is performed, and bit transition among concatenated zero crossings is counted at corresponding locations to compute Sequency plane, Seq. Finally, Seq is encoded into Sequency code bit plane feature SeqC. Because of its simple fusion strategy, the Sequency code requires less computations and time in feature extraction process. Performance of Sequency code is validated by carrying out large number of experiments with three different standard databases which are PolyU 2D, PolyU 2nd version, IITD and Multispectral databases. [7]

Liliana Studied about biometrics of palm for identification system using the physical form of human hands, represented by the image of the palm. The methodology for palm recognition includes 2 parts.

I. Block-based line detection which dealing with palm print feature extraction process.

II. II. Chain code which solved the land geometric feature extraction.

These 2 respective features were combined and it was introduced as Dynamic Time Warping (DTW) method. It

can measure the distance between two different features. Then it was proposed that the result can be used for personal identification in authentication systems for information security.[9]

Inspecting the above details and compare with the other biometric techniques, the palm print is the relatively new biometric feature and is regarded as one of the most stable, reliable and unique personal characteristics. So the main purpose of this research project is to provide a Fast and high Accuracy Palm print Recognition system for low-quality palm-print images using below methodology.

II. METHODOLOGY

The block diagram of the proposed method architecture has been shown in following and this method architecture can be divided into three main steps.

1. Image preprocessing
2. Feature extraction
3. Palm-print matching

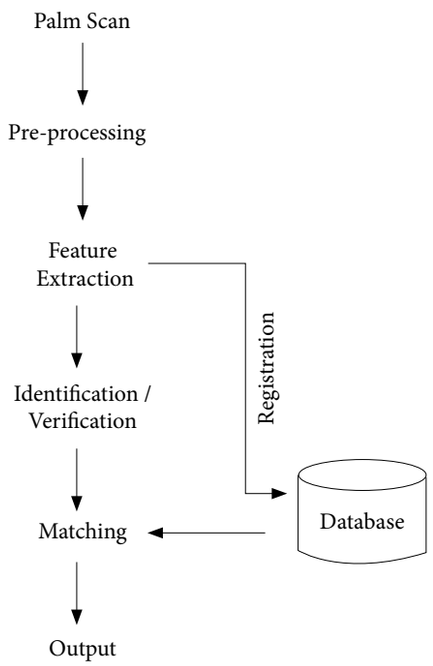


Figure 1. Proposed method architecture

A. Pre-processing, Database, and Region of Interest (ROI) Extraction

Image preprocessing is the first step in pattern recognition. There are two objectives in preprocessing.

I. Obtaining a sub palm-print image for feature extraction.

II. Eliminating the variation happens due to rotation and translation.

This preprocessing block is one of the most critical part of the developed palm-print recognition algorithm, and it is the first block in the developed algorithm as it is the case in many biometric systems. Before feature extraction and coding step, two things must be done.

a. Preprocessing all images in the database.

b. Obtaining the central area of a palm.

The main purpose of this pre-processing step is to extract the certain region form the palm-print which includes principal lines, ridges and wrinkles. Extracting is done by compensating for rotation and translation. The correct extraction of ROI plays a crucial role in improving the performance of the overall palm-print recognition. Indian Institute of Technology, Delhi already provided an ROI image of the database. The IIT Delhi Touchless Palm-print Database version 1.0 has been used during the development of this research and this database mainly consists of the hand images collected from the students and staff at IIT Delhi, India.

B. Feature Extraction

A palm-print can be represented by some line features from low-resolution images. Algorithms such as stack filter are able to extract the principal lines. But, these principal lines are not enough to represent the uniqueness of each individual's palm-print as different people can have same principal lines in their palm-prints. And also some palm-print images do not have clear wrinkles. To extract features from palm-print image several techniques have been implemented such as wavelet. But these methods are unable to detect clear edges and smooth curves which result from the conjunctions of principal lines and wrinkles. In this research project, extract features from

low-resolution palm-print images by using IIT Delhi Touchless Palm-print database version 1.0. Following image shows the sample ROI of a palm-print image from the above-mentioned database for extracting principal lines. The software which has been used for this research project was the Matlab.

Noises of the image should be removed before extracting palm lines from ROI image. For this step Wiener 2 lowpass filter is used, shows in figure 3. It is a grayscale image which has been degraded by constant power additive noise. Wiener2 uses a pixel-wise adaptive Wiener method based on statistics estimated from a local neighbourhood of each pixel. This approach produces better results than linear filtering.

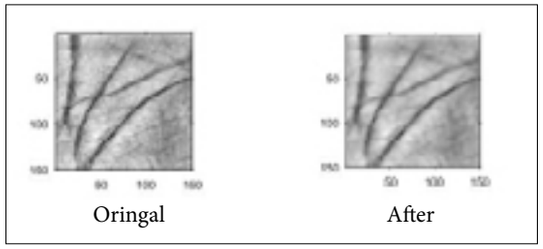


Figure 2. applying the lowpass

After this step, elongated and tubular structures can be enhanced in the noise-removed image for highlights major lines by using Hessian-based multi-scale filtering. The returned image contains maximum response of the filter at a thickness which approximately matches the size of the tubular structure in the image. Figure 4 shows the enhanced version of the noise removed image which highlights threads that are seven pixels thick.

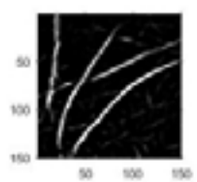


Figure 3. An enhanced version of the image that highlights threads that are seven pixels

Usually, palm-print identification methods rely on grayscale pictures and samples. It is because colour identification is not essential to identify a subject. After completing the segmentation process the original image has to be transformed into a binary image in which ridges area fully colored in one tone and the background in the opposite tone. Image binarization is fundamental to find principal lines or other singularities.

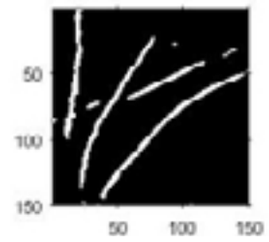


Figure 4. Threshold result

And this image binarization is done in other computer version processes as well as palm-print identification. Image thresholding is ineffective way to partition an image into a foreground and background. It isolates objects of the image by converting grayscale images into binary images. The binary image is specified as a non-sparse, local or numeric array of any dimension. This image thresholding is most effective in images with high levels of contrast. In this threshold image, as shown in figure 4, each pixel represents binary values 0 and 1. Finally, Threshold binary image is applied to some morphological operations for extracting better results.

III. MATCHING PALMPRINTS

After completing this line extraction process, features are extracted for matching. Here, 50 palm print images have been selected randomly from 10 different persons and each person have 5 different palm images. In this palm-print matching step, one palm-print is selected from each person as the template image. This template image is subjected to the principle line extraction process. Then those specific output values are enrolled which were obtained from that template image. Total 10 enrollments are saved for the matching. In this step, the palm-print matching process is based on normalized cross correlation in Fourier domain and it is also known as phase correlation. Here in this context, correlation can be called as template matching and the two variables are the corresponding pixel values in two images, template, and source.

Cross-correlation is a measure of the displacement of one relative to the other. In normalized cross-correlation, two

images are used to snapshots of the same scene. Template matching techniques compare portions of images against one another and sample images are used to recognize similar objects in the source image. Template matching is used when the standard deviation of the template image compared to the source image is small. This matching process moves the template image to all possible portions in a larger source image and a numerical index is computed which indicates how well the template matches the image in that position. This matching process is done on a pixel-by-pixel basis.

A. Cross-Correlation Analysis

Cross-correlation is a kind of template matching, between two input signals. It can be done in any number of dimensions. One-dimensional normalized cross-correlation between two input signals can be defined as follows.

$$r_d = \frac{\sum_i [(x[i] - \bar{x}) \cdot (y[i-d] - \bar{y})]}{\sqrt{\sum_i (x[i] - \bar{x})^2} \sqrt{\sum_i (y[i] - \bar{y})^2}}$$

r – a measurement of the size and direction of the direction of the linear relationship between variables x and y. r= +1, if these variables move together, where they both rise at an identical rate. r= 0, if the other variables do not budge. r= -1, if the other variable falls at an identical rate. If r is greater than zero, the correlation is positive and r is less than zero the correlation is negative. The sample non-normalized cross-correlation of two input signals requires that r be computed by a sample-shift (time-shifting) along one of the input signals. For the numerator, this is called a sliding dot product or sliding inner product.

The dot product is given by the following equation.

$$X \cdot Y = \sum_i x_i y_i$$

When equation 1 is computed, for all delays, then the output is twice that of the input. When showing a cross-correlation, usually the pentagon notation can be used.

$$(X \star Y)_d = \sum_i x_i y_{i+d}$$

Where the asterisk indicates the complex conjugate which is a negation of the imaginary part of the number. Input signals should either have the same length or a policy in place to make them the same (perhaps by zero padding or data replication). Following equation can be written if the input signals are real-valued.

$$(x \star y)_d = \sum_{i=-\infty}^{\infty} x_i y_{i+d}$$

Comparing equation 4 with the convolution.

$$x_n * y_n = \sum_{i=-\infty}^{\infty} x_i y_{n-i}$$

Y is time-reversed before shifting by n. In comparison, the correlation has to shift without the time reversal.

B. Fourier Transform Analysis

The Fourier transform is a representation of an image as a sum of complex exponentials of varying magnitudes, frequencies and phases. It plays a critical role in a broad range of image processing applications including enhancement, analysis, restoration, and compression.

If f(m, n) is a function of two discrete spatial variables m and n, then the two-dimensional Fourier transform of f (m, n) is defined by the following relationship.

$$F(w_1, w_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m,n) e^{-iw_1 m} e^{-iw_2 n}$$

F (ω_1, ω_2) – Frequency-domain representation of f (m, n).

It is a complex-valued function that is periodic both in ω_1 and ω_2, with period 2π. Because of the periodicity, usually, only the range – π ≤ ω1, ω2 ≤ π is displayed. Note that F (0, 0) is the sum of all the values of f (m, n). For this reason, F (0, 0) is often called the constant component or DC component of the Fourier transform. (DC stands for direct current; it is an electrical engineering term that refers to a constant-voltage power source, as opposed to a power source whose voltage varies sinusoidally).

The inverse of a transform is an operation that when performed on a transformed image produces the original image. The inverse two-dimensional Fourier transform is given by

$$f(m,n) = \frac{1}{4\pi^2} \int_{w_1=-\pi}^{\pi} \int_{w_2=-\pi}^{\pi} F(w_1, w_2) e^{-iw_1 m} e^{-iw_2 n} dw_1 dw_2$$

IV. RESULT AND DISCUSSION

In this research project the palm recognition phase has been divided into two stages. The first one is called intra-class comparison. In this study, only one was rejected out of 50 comparisons. This rejection called as false rejections. To check inter-class calculation there are 450 comparisons. Out of these 450 comparisons, there were 44 wrongly accepted results shown. This results called false acceptances.

The performance of biometric security systems can be evaluated from biometric authentication parameters. Those parameters are false accept rate, false reject rate, genuine accept rate, accuracy, half total error rate etc.

False Accept Rate (FAR) is the percentage of faulty/ incorrectly recognized users. FAR is a measurement which explains the percentage of faulty recognitions of unauthorized individuals.

$$FAR = \frac{No.of\ false\ acceptance\ found}{Total\ No.of\ Comparisons} \times 100$$

In this study, the calculated FAR is 8.8% according to above experimental results. 44 had been wrongly accepted out of 450 comparisons.

False Reject Rate (FRR) is the measure of the incorrectly rejections of access attempts by authorized individuals, of biometric security system. The FRR is calculated as,

$$FRR = \frac{No.of\ false\ rejection\ found}{Total\ No.of\ Comparisons} \times 100$$

In this study, the calculated FRR 0.5%. One was rejected from 50 comparisons.

Genuine acceptance rate (GAR) is the percentage of genuine users who were accepted by the system and it is defined as,

$$GAR = 1 - FRR$$

In this study, GAR is 99.8%. The GAR should be high.

Also Half Total Error Rate (HTER) is a possible way to measure the performance of the biometric system. It combines both types of system errors of false accept and false reject. It is defined by following formula:

$$HTER = \frac{1}{2}(FAR + FRR)$$

In this study HTER is 4.5%.

Accuracy of the system is the percentage efficiency of the system in terms of its ability of authentication. The accuracy is calculated as,

$$Acuracy = 100 - \frac{FAR + FRR}{2}$$

In the study the calculated accuracy of the system is 95.5%.

Furthermore, the proposed algorithm is implemented using MATLAB R2017a image processing toolbox. The execution time for the feature extraction and coding block is less considerable. Finally, the execution time for the template matching is about 20 milliseconds. Since total verification time does not depend on the number of templates in the database, it can be found by summing up feature extraction and coding block and template matching block. Therefore; the total verification time is calculated to be around 1 second, which is acceptable for a high accuracy palm-print recognition algorithm.

V. CONCLUSION

This research has been conducted in order to develop a Fast and Accuracy Palm Print Recognition System for Low-quality Patterns. Research questions have been answered so that the research objectives are met. It has been identified that most biometric security system studies have been done using finger prints, facial recognition, voice recognition or iris recognition. And there are very fewer number of studies have been identified which has been done using palm prints as the input for the biometric security system. When considering about this proposed algorithm which includes principle line feature with template matching, performances of it is suitable for nowadays commercial applications.

ACKNOWLEDGEMENT

I offer my sincerest gratitude towards to my supervisor Mr. L. R. Dangalla, lecturer of The Department of Computing and Information System, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. This research and thesis would not be possible without the help of him and he always guided me to solve the problems which arise throughout this research. I am extremely thankful and grateful to him.

And also I would like to express my thanks to Head of the Department of Computing and Information Systems Dr. B.T.G.S.Kumara, who guided me throughout the research period giving valuable advices.

REFERENCES

Babich, A., 2012. Biometric Authentication. Types of biometric identifiers.

Bubeck, D.S.U. and Sanchez, D., 2003. Biometric Authentication. *Universidade Estadual de San Diego*.

Singh, S., Correia, P.L. and Soares, L.D., 2012. Partial Palmprint Matching for Forensic Applications. *M. Tech Thesis*.

Raghavendra, R. and Busch, C., 2015. Texture based features for robust palmprint recognition: a comparative study. *EURASIP Journal on Information Security, 2015(1)*, p.5.

Logapreethi, M. and Kumar, N.M., A Fast and Accurate Palmprint Identification System based on Consistency Orientation Pattern.

Khokher, R., Singh, R.C. and Kumar, R., 2014, March. Palmprint recognition using geometrical and texture properties. In 2nd *International Conference on Research in Science, Engineering and Technology (ICRSET'2014), Dubai (UAE)*.

Dubey, P., Kanumuri, T. and Vyas, R., 2018. Sequency codes for palmprint recognition. *Signal, Image and Video Processing, 12(4)*, pp.677-684.

Zia-uddin, Z.J., Jan, Z., Ahmad, J. and Abbasi, A., 2014. A novel technique for principal lines extraction in palmprint using morphological Top-Hat filtering. *World Applied Sciences Journal, 31(12)*, pp.2010-2014.

Liliana, D.Y. and Utaminingsih, E.T., 2012. The combination of palm print and hand geometry for biometrics palm recognition. *International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS, 12(01)*.

Jia, W., Hu, R.X., Gui, J., Zhao, Y. and Ren, X.M., 2012. Palmprint recognition across different devices. *Sensors, 12(6)*, pp.7938-7964.

George, A., Karthick, G. and Harikumar, R., 2014, March. An efficient system for palm print recognition using ridges. In *Intelligent Computing Applications (ICICA), 2014 International Conference on* (pp. 249-253). IEEE.

Manegopale, P., 2014. A Survey on Palm print Recognition. *International Journal of Innovative Research in Science, Engineering and Technology, 3(2),* pp.9085-9094.

Raut, S.D. and Humbe, V.T., 2012. Biometric palm prints feature matching for person identification. *International Journal of Modern Education and Computer Science, 4(11)*, p.61.

Salave, H. and Pable, S., 2015. Improved palmprint identification system. *International journal of scientific & technology research, 43*, pp.180-185.

Lyon, D., 2010. The Discrete Fourier Transform, Part 6: Cross-Correlation. *Journal of object technology, 9(2)*, pp.17-22.

SMART TEA LEAVES DISEASE ANALYSER: MOBILE BASED DISEASE DETECTING AND SOLUTIONS PROVIDING SYSTEM

[1]VS Dahanayake, WJ Samaraweera and DMR Kulasekara
Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka
[1]*vasitha.sd@gmail.com*

Abstract - Sri Lanka is well-known for its excellent TEA and as the 3rd largest tea manufacturing nation internationally. Sri Lanka is one of the main world's top TEA exporters with a high global demand attracting millions of foreign exchanges, which strengthens the backbone of the economy of the country. Tea is grown in the whole country including central highlands and southern inland areas, resulting in a lot of diversity in the taste of Sri Lankan tea. Although TEA is an important agricultural field, the lack of attention, lack of resources, high cost of production has reduced its productivity and quality. One of the main reasons for this low productivity can be identified as tea leaf diseases due to changes in weather, infertile soil, pests, etc. This research paper suggests an automated and economical methodology to draw-up the current inefficient manual process of disease detection of tea leaves in tea cultivation by new trends of computing field such as image processing and machine learning techniques. The steps such as Image Acquisition, Image Segmentation, Image Pre-processing, Feature Exaction and Classification and Detection of tea leaf diseases are developed into an android application to provide effective, efficient, cost-effective and highly accurate system which will establish safer growing conditions. Also, the suggested solution will reduce the environmental and ecological impact due to usage of chemicals only in necessity to approved amounts to recommendations of the mobile application.

Keywords- image processing, leaf disease detection, mobile technology, smart agriculture

I. INTRODUCTION

Tea plantation is a main source that brings foreign exchange to Sri Lanka, Sri Lanka produces 2% of Gross domestic product (GDP) of tea, donating billions to the economy of Sri Lanka. (EDB, 2014) Today, there is a high demand for Ceylon tea both locally and internationally. However, Ceylon tea plantation is endangered due to various plantation diseases occurring as a result of different environmental circumstances and Sri Lanka's farming efficiency stays low contrasted with different countries at comparable levels of advancement. (Oxford business group, 2018) Due to environmental changes such as rainfall, temperature, pests, the crop yield gets affected severely. Development of automatic detection system of tea leaf diseases using advanced computer technology such as image processing helps to support the farmers in the identification of diseases at an early or initial stage and provides useful information for its control. Therefore, the present study was carried out on automatic disease identification of tea leaf using image processing and machine learning techniques, which includes Image Acquisition, Image Segmentation, Image Pre-processing, Feature Exaction and Classification and Detection of tea leaf Diseases.

With regard to the development of technology people try to make things easier. Considering this situation, it is important to develop a method to facilitate the issues with

agricultural sector, tea plantations. To overcome those issues of tea leaves(diseases) in the current agricultural field, An Android Application, which can be used as a self-usable tool can be produced.

Over a decade, the mobile technology, such as smart-phones, tablets, different applications, etc. has become a requisite part of the human life in all over the world. With the time development of the mobile technology, the uses and the services of the mobile device have increased. The mobile technology went beyond the traditional telecommunication. It changed the way of communicating, sharing information and how their works are done. With that development, many other fields such as education, agriculture, healthcare, business, also developed.

Likewise, with the combination of mobile technology and the generation's new cognitive new solutions for the current concepts on agriculture have changed dramatically. As one prospective solution this paper suggests an Android based mobile application that detects the diseases at an early or initial stage and provides useful information for its control or extinction.

Whenever, user is having an issue with tea leaves they can input those images to the application using its image recognizable scanner. Then it will provide the solution with all the necessary steps. It will also provide appropriate diagrams and other similar solutions.

Because of this proposed system, farmers do not want to refer many reference books and online tutorials or wish for the assistance of tea inspectors. They can manage to find the solution using their fingertips. It will reduce the high time consumption, that they have spent on waiting.

The proposed system is a, system which contains the graphical symptoms of tea leaves and other parts of the tea plant, so that the users will be able to use their mobile device whenever necessary to scan the relative infected leaf and diagnose the disease in a much effective, less time consuming and a very much cost-effective way

A study done by (Deshmukh, n.d.) states that professional detection of plant diseases can be excessively high in cost through naked eyes, specifically in developing countries. As a solution paper discusses providing a fast, programmed, cost-effective and precise solution which makes the image processing as the base. For the leaf disease detection field, it can be a great idea which could be developed using relevant techniques.

Smart mobile phone has become an essential thing for the people today. Therefore, using an application like this would be much more useful than using human assistance. This solution will be user-friendly and will perform in an effective and efficient manner.

The importance of this system is mainly to the tea inspectors and farmers whose field is tea cultivations. This system will identify and categorize the infected leaves and will provide solutions to the diseases. It will be an effective, less time consuming and a very much cost-effective way.

Mainly the system will Greater efficiencies and lower prices of the tea cultivations, Safer growing conditions will be established and the reduced environmental and ecological impact due to usage of chemicals only in necessity to approved amounts to recommendations of the app. The main significance is that this provides solutions to inefficient process of disease detection of tea leaves in tea cultivations.

The paper is incorporating with several sections. Section 2 is the literature survey for the topic; Section 3 presents the research findings and Section 4 describes the conclusion along with the references.

II. LITERATURE REVIEW

The old and classical naked eye observation-based tea leaf disease detection systems are very inefficient, slow and also give less accuracy. And also consulting experts to find out plant disease is expensive and time consuming due to unavailability of experts.

Tea leaf disease is one of the critical reasons that decreases quantity and degrades value of the tea. (Karmokar et al., 2015) at his study has shown that the leaf disease of a plant can interrupt the normal functions of the plants like photosynthesis, fertilization, growth, transpiration etc. badly. The main indicator of a leaf disease is the colour of the tea leaf, they change colour when the leaves are diseased. Also, the study shows that the plant leaves show normal colour until a plant leaf get affected by any sort of pathogen after that it changes gradually.

Review paper by (Raut and Ingole, 2017) states the basic steps of image processing is image storing and keeping them for the future use.

In this study conducted by (Qin et al., 2016), observes and identifies four types of alfalfa leaf diseases using image-processing techniques and uses pattern recognition algorithms as the base.

As in the many cases (Chaudhary et al., 2012) has identified the plant diseases using computer vision. In this study the diseases were detected by extracting colour feature. The colour models they have used are YcbCr, HSI, and CIELB. They were successfully able to detect diseases without the noises by different sources such as camera flash.

Although there are many plant disease detection methods via automation or computer vision there are no marketable resolutions found according to the paper by (Sladojevic et al., 2016). They also have found a new method of using deep learning which decides and detects plant diseases from leaf images by itself. The established prototype was able to detect leaf presence and differentiate between healthy leaves and 13 different diseases successfully.

In the paper (Padmavathi and Thangadurai, 2016) Image Pre-processing has been identified as an effective and consistent way to identify leaf diseases which includes an assortment of methods which are used in progress of improving graphical presence of an image. The extraction of colour and texture feature extraction in the leaves are the most significant factors according to the authors. According to this paper, diseases as they have given noise free images which are well suitable for human or machine understanding when compared with grayscale colour images.

(Arivazhagan et al., n.d.) has created processing system that consists of four basic stages, where colour transformation structure for the input RGB image is created as the first step, After the segmentation process as the second step the green pixels are masked and removed, the texture numbers are added for the beneficial parts, As the last step a classifier is used to pass the mined features.

(Varshney and Dalal, 2016) states that some of the most popular plant leaf sorting techniques are, Genetic Algorithm, Neural Network, Principal Component Analysis, Support Vector Machine and k-Nearest Neighbour Classifier. In the process of plant leaf classification most important factor that the paper identifies is the morphological features in the plant leaf.

The study by (Deshmukh, n.d.) shows that usage of K means clustering and neural network which is used for segmentation and classification of diseases pave the way to detect the leaf disease accurately and speedily. They have conducted the study around the paddy leaves and the approach presented in the paper is image processing based.

A paper presented by (Al-Hiary et al., 2011) proposes an algorithm containing the following eleven steps for plant disease detection which uses techniques such as K-means clustering, Masking, RGB to HIS Translation, Neural Networks for Recognition.

Also (K. Singh and Chetia, 2017) has developed a working software on Identification and Arrangement of Plant Leaf Diseases in Image Processing using MATLAB. The paper discuses on automatic disease detection of "Phaseolus vulgaris" and "Camellia assamica" plant leaf using image processing techniques.

III. METHODOLOGY

The complete process of finding the model for tea leaf disease recognition from many literatures and the interview is described further in detail below. The complete procedure is divided into some essential steps in subsections below;

A. Data Gathering

Data that is essential to conduct the research were collected through an interview and literature and article reviews. The interview was directed with a tea inspector to get the full idea on the manual leaf disease detection process.

C. Data Analysing

The data which was gathered during the data collection process was analysed in this phase to define the problem and to identify the limitations with the existing process.

Suggestions of the users to improve the current process could be identified through the collected data. Consuming much time and cost, limited number of resources, lack of experience and training to the tea inspectors are some of the drawbacks with the manual procedure which could be identified through data analysis.

D. Approach

Users of this system are farmers, tea inspectors and tea. Inputs for the system are the images of diseased tea leaves taken by the user of the app using android mobile phone. Outputs from the system are basically a report that states the matching disease and available remedies for the relevant disease and the further steps that could be taken to future spreading of the disease or the bug. The system receives inputs and executes user requests to generate a report containing the disease information and remedies.

E. Technology adopted

The developed system consists of a mobile application. The mobile application has been developed using MATLAB Simulink which is developed by MathWorks, where the design signal or image processing algorithms and applications for Android devices are supported. Image enhancement process was designed through image processing filters and functions in MATLAB Simulink. An Artificial Neural Network that is known as a supervised machine learning prototype which identifies the patterns based on a previously trained dataset, has been integrated in order to develop this disease detection and providing solution on the basis of the disease.

As for the development platform of database SQLite will be used. The main reasons to use SQLite are Zero-Configuration, Serverless, Stable Cross-Platform Database File. Which will be most suitable as the developing solution is an android based application

F. Design

The overall architecture of the system can be defined based on three main layers; client layer, application layer and database layer.

High level architectural perspective of the developed system is shown in figure 1 below.

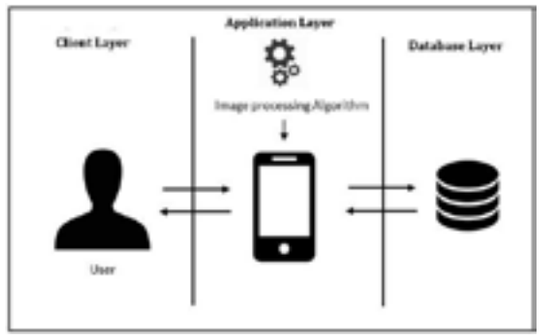


Figure 1. High Level System Architecture

1) Client Layer: The client layer will receive input from the Interface modules. This layer allows the admission to the users of this system. The two main user levels are tea inspectors and farmers. Access levels vary from one user level to the other in the system due to the technical knowledge barrier between the two different users. In this section, the authorized users will be predefined.

2) Application Layer: The application layer contains an android mobile application. This is to be further described through the below discussed modules.

   i. Login Module
   This module confirms the username and password that are entered by the clients and gives the entrance to the system for the predefined approved clients. It has given an option to recuperate username and password if forgotten.

   ii. Insert Information Module
   In this module Authorized Tea Inspectors can provide the information on leaf diseases and images to the system. Then a verification process of the inserted information is done to make sure that the entered details are accurate before adding them to the system.

   iii. Update Information Module
   Only authorized users are allowed to update information on diseases. Then a verification process of the inserted information is done to make sure that the entered details are accurate before the changes are applied to the system.

   iv. Delete Information Module
   Only authorized users are allowed to delete the information on diseases. Then a verification process of the deleted information is done to make sure that the entered details are accurate.

   v. The Identical Disease Matching Module
   The "image enhancement" option has been provided to improve the quality of the image. Then the enhanced image is used to search against the database to find a full match or a partial match for the disease. A report will be generated indicating the disease information, suggested remedies and matching probability and other relevant details.

   vi. Email Sending Module
   The generated reports will be sent to the relevant personals through emails in this module.

   vii. SMS Sending Module
   This module allows to send SMS alerts automatically to inform about the emails.

3) Database Layer: The Database Layer is accountable for handling the database of the system. The database will keep the data that are given to the system by the applications. And also, this layer is responsible to retrieve the necessary data when requested.

Primary steps for tea leaf disease detection and identification can be mentioned as follows;

G. Image Acquisition

Diseased tea leaf images were captured with a mobile phone camera and stored for the future use in MATLAB Simulink in RGB format.

Figure 2. Diseased tea leaf from 'Blister blight' (A plant pathogen)

H. Image Pre-Processing

The image Pre- processing is performed to enhance the quality of the current RGB image and to remove the noise in the image. Image is converted to greyscale from RGB. This task is mainly performed by clipping and smoothing the image. Also, the increasing contrast of the image is done by image enhancement.

I. Image Segmentation

Image segmentation is used to alter the digital image into different segments which look alike. This method helps in identification of borderline of the leaf.

J. Feature Exaction

This is the step where all the facts texture, colour, breed and arrangement of the leaf are considered. Each and every simple detail is important to the disease detection process by using considering similar occurring colour methods features and diseases are separated.

K. Classification and Detection of Tea Leaf Disease

As in the above step features are separated and, in this step, the retrieved diseased leaf parts are matched with the database and the infected disease is provided by the database for the closest match.

IV. RESULTS AND DISCUSSION

The previously developed systems with reference to the literatures have standalone applications mostly. In the practical usage of these systems the users have to scan and photograph the diseased leaf using a digital camera and then upload it to the application. In this system the android application allows to scan the infected tea leaf using the app itself, so this application is more convenient, fast and accurate to use.

The above identified novel system follows primary steps for tea leaf disease detection and identification as Image Acquisition, Image Segmentation, Image Pre-processing, Feature Exaction and Classification and Detection of tea leaf Diseases. The importance of this system is mainly to the tea inspectors and farmers whose field is tea

cultivations. This system will identify and categorize the infected leaves and will provide solutions to the diseases. It will be an effective, less time consuming and a very much cost-effective way. Mainly the system will Greater efficiencies and lower prices of the tea cultivations, Safer growing conditions will be established and the reduced environmental and ecological impact due to usage of chemicals only in necessity to approved amounts to recommendations of the app. The main significance is that this provides solutions to inefficient process of disease detection of tea leaves in tea cultivations.

The main objectives of the application are Identifying diseases in tea cultivation before crop yield gets affected severely, reducing the time wastage in finding the disease and necessary remedies and the make the tea leaf disease detection process more accurate and reliable, reducing the waiting time for tea inspector's expertise knowledge, and to make the complex process of identifying disease process simple easy and convenient.

The issues and challenges of the system can be identified as the technology Upgrades which is one of the biggest concerns of developing the solution as technology is growing and developing fast day by day. Also, as the farmers and general public are the targeted audience of the application, the technological knowledge barriers will arise when using the solution. Needing a large trained dataset to identify diseases on the basis of the previously identified patterns is another challenge. Also, failure to capture the images of the leaves to the required quality can also be identified as an issue of the application.

V. CONCLUSION AND FURTHER WORK

In this research work, the authors could develop an efficient and accurate Tea Leaf Disease Detection System by integrating novel advancements in computer science field in order to speed up the traditional Tea Leaf Disease Detection process. It will help in quick and accurate identification of diseases that leads to enhance the efficiencies and lower prices of the tea cultivations where safer growing conditions will be established and the reduced environmental and ecological impact due to usage of chemicals only in necessity to approved amounts to recommendations of the app. The main significance of the newest system will be that this provides solutions to inefficient process of disease detection of tea leaves in tea cultivations.

The images taken by mobile phone cameras may be different in size, focus, quality and also the distance between the leaf and the mobile phone can vary from person to person even though there are standards that are established. For this the authors infer to use a drone to take pictures in the tea estate section wise from a certain distance and analyse the images for the diseases in the whole area for the further work. This will change the whole idea on disease detection process on tea leaf. By analysing the tea leaves in the drone photographs the users will able to check the amount of sunlight, fertilizer and water that the tea plant gets. Finding a novel, accurate and efficient mobile solution for tea leaf disease detection system using drones will be an interesting future direction that will help to lead a great upheaval agriculture and export fields in Sri Lanka.

REFERENCES

Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., ALRahamneh, Z., 2011. Fast and accurate detection and classification of plant diseases. Mach. Learn. 14.

Arivazhagan, S., Shebiah, R.N., Ananthi, S., Varthini, S.V., n.d. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features 15, 7.

Chaudhary, P., Chaudhari, A.K., Cheeran, A.N., Godara, S., 2012. Colourtransform based approach for disease spot detection on plant leaf. Int. J. Comput. Sci. Telecommun. 3, 65–70.

Deshmukh, R., n.d. Detection of Paddy Leaf Diseases. Energy Angular Second Moment 2, 4–3.

K. Singh, M., Chetia, S., 2017. Detection and Classification of Plant Leaf Diseases in Image Processing using MATLAB. International Journal of Life Sciences Research 5, 120–124.

Karmokar, B.C., Ullah, M.S., Siddiquee, M.K., Alam, K.M.R., 2015. Tea leaf diseases recognition using neural network ensemble. Int. J. Comput. Appl. 114.

Oxford business group, 2018, Increasing agricultural production and exports in Sri Lanka are key goals for sector, viewed 20 February 2018, <https://oxfordbusinessgroup.com/overview/cream-crop-increasing-both-production-and-exports-are-key-goals>.

Padmavathi, K., Thangadurai, K., 2016. Implementation of RGB and Grayscale Images in Plant Leaves Disease Detection – Comparative Study. Indian J. Sci. Technol. 9.

Qin, F., Liu, D., Sun, B., Ruan, L., Ma, Z., Wang, H., 2016. Identification of Alfalfa Leaf Diseases Using Image Recognition Technology. PLOS ONE 11, e0168274. https://doi.org/10.1371/journal.pone.0168274

Raut, S., Ingole, K., 2017. Review On Leaf Disease Detection Using Image Processing Techniques.

Sri Lanka Export Development Board, EDB, 2014, INDUSTRY CAPABILITY REPORT, viewed 18 February 2018, <http://www.srilankabusiness.com/pdf/industrycapabilityreport_tea_sector.pdf >.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification [WWW Document]. Comput. Intell. Neurosci. https://doi.org/10.1155/2016/3289801

Varshney, S., Dalal, T., 2016. A Novel Approach for the Detection of Plant Diseases.

HAIRSTYLE RECOMMENDATION BASED ON FACE SHAPE USING IMAGE PROCESSING

SV Rajapaksha[1] and BTGS Kumara
Department of Computing & Information Systems, Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, Belihuloya
[1]*shashikavidr@gmail.com*

Abstract- Hairstyling is an art of fashion transformed since ancient era, with the influences from many diverse factors. It has been a primary aspect of human lifestyle and society in various different ways with the growth of research fields like modelling human, visual searching, visual matching, facial verification for security measures and etc. Perfect hairstyle improves specially a woman's self-confidence. This paper presents a hairstyle recommendation system based on face shapes and suitable hairstyle with expert's knowledge for the face shape derived from face shape classification algorithm. Recommendation algorithm has developed base on the learning relationship between facial shapes and suitable hairstyles. This research has classified face shapes into 5 different shapes: round, oval, oblong, square, and heart. Here, machine learning libraries were used to detect the landmarks of a face image in face shape identification process. The accuracy of our face shape identification algorithm is 85% out of 100 images. After identifying the shape of the face, the recommendation system proposes suitable hairstyles for the face image. Here, have used Python programming language and image processing techniques to develop the algorithm. The system will allow users to upload a preferred face image, process it and will automatically select the matching hairstyles category for the given image. The empirical study of our prototyping system has proved the effectiveness of our recommendation algorithm.

Keywords- Face Shape, Facial Feature, Hair Styling, Image Processing, Landmark Detection

I. INTRODUCTION

The hairstyle is one of the most important aspects of people in determining their appearance and mood. People look completely different by changing their hairstyles. Most one deeply cares about their hairstyles not only women but also men's. The Hairstyle can make human appearance attractive or unattractive. If someone chooses an inappropriate hairstyle, then it gives a bad looking and loses confidence.

A hairstyle mean is styling hair on the human scalp. Hair gives various fashionable styles to a human's body. Which is main aspects of the human body although without hair become an unnatural but also without hair is a new hairstyle to humans. The increase of fashions most people think of hair as their main important thing.

Many people prefer to choose their hairstyles from magazines, internet without knowing which one is realty suits their faces. If woman's hair cut too short, it takes a long time to enough grow for doing new hairstyle. Therefore, before going to the salon it will better choose suits hairstyle to our face. Sometimes our selected hair styles do not match to our face. A beauty expert says that a proper hairstyle for someone depended on their face shapes. It is better to know our face shape and features well before doing hairstyles. Similar face shapes have similar hairstyles. Therefore it is better to have hairstyles recommendation system to know about hair styles before doing hairstyles. A face shape classification is essential for the development of a hairstyles recommendation system.

Several commercials software has been developed for allowing users to simulate how they look with different hairstyles by manually changing. Other researches use face modelling techniques for face detection and face recognition works. The major objective of the work is a proposed method to recommend the hairstyles based on major face shapes with a combination of hair expert's knowledge. One of significance in the proposed methodology mainly concern on image processing techniques to detect face shape rather than using other AI techniques. The proposed classification algorithm is to classify the face shapes into five shapes oval, oblong, square, round and hart.

II. METHODOLOGY AND EXPERIMENTAL DESIGN

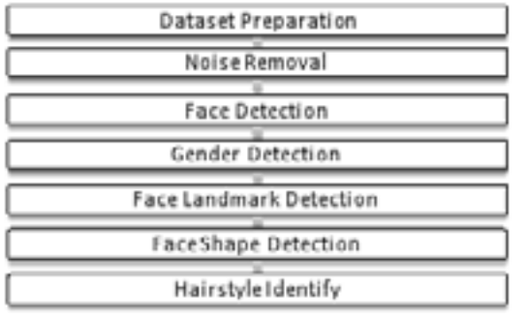


Figure 1. Overview of methodology

A. Data Collection

Collected information about hairstyles, hair features and face shape from magazines, websites and person who expert about the hair. Mainly target the salons in the various places in Sri Lanka for collet the hairstyles information and other facial features. Basically, salon owners think about customers and event type before making hairstyles. Both male and female features and hair styles are different. We have chosen face shape and facial features information to identify the matching hairstyles.

We have collected a large face dataset that contains hairstyles images under the unique categories such as hair types (straight, wavy, curly) and further categorized as male and female as well. Each of those categories we collected more than 1000 images. There are five face shapes, rounded, oval, and squared and heart.

B. Face Detection

To identify the faces we use the Haar Cascade classifier in the OpenCV. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

C. Face Landmark Detection

Face Landmark detection is most the important part when it comes to face shape identify. In order to detect face landmarks, the Dlib machine learning library was used, which is written in the C++ library. The OpenCV/C++ library was used for python, there is an API support for python.

Features of all images were extracted by landmark detection. Figure 2 shows the identified eyebrows, eyes, mouth, chin and all 67 points in the face. But, there were no any defined coordinate points which separate a forehead and the hair. Researching about face, it would help to find other points in the face. After finding all points in the face, we have generated the algorithm to find the face shapes.

To identify the Figure 3 shown points, the researchers have used thresholding techniques to separate the forehead from the hair. Before use thresholding techniques it is better to make an image as a grayscale image. The input image consists of with three color channels. Where contains Red, Green and Blue channels. It's easy to do threshold processing by converting those three channels into a single channel. In order to make that happen here, we used a gray scaling method in OpenCV.

After converting the image into a grayscale image, There have tried on simple thresholding methods like Binary, Binary inverse thresholding, To Zero thresholding likewise. The most complex task was to the identify threshold value because the documents have different colors and when the scanning, pixel luminance may change. To overcome the problem research use the

adaptive thresholding Otsu method. Since researches are trying to make binomial images along the process we are doing and it is the best input to Otsu thresholding because of Otsu thresholding algorithm find the Bimodal values and try to find the threshold that minimizes the weighted within-class variance.

D. Identify Face Shapes

Figure 2. facial landmark points identified on face

In order to find the face shape, there are several calculations have done. Table 1 denotes what are the parameters take part in the proposed face classification calculations.

Figure 3. The extra coordinate values which need to identify

Table 2 shows the face shapes and its condition with face length and angles. In order to make those conditions the knowledge gain from previous researches as well as experts in the industry and other studies related to face modelling is used.

E. Select the Hairstyles

The hairstyles are categorized for relevant face shapes. Face Shapes have different hairstyles. According to the expert's knowledge hairstyles are stored in the database base on the face shapes. After identifying the face shapes, the system finds the matching hairstyle from the database.

Table 1. Calculated lengths and angles on image

Coordinate points	Description
68-8	Face Length
70-69	Forehead Length
8-3	Jawline Length
10-6	Chin Width
26-17	Cheekbone Width
1-4	Angle a1
4-6	Angle a2
6-8	Angle a3

Table 2. Condition for the face shapes

Face Shape	Condition
Heart	Forehead Width > Cheekbone Width > Jawline && a1>a2>a3
Oblong	Face Length > (Cheekbone Width ≈ Forehead Width ≈ Jawline) && a1 ≈ 90>a2>a3
Oval	Face Length > Cheekbone Width & Forehead Width > Jawline && a1 ≈a2≈a3
Square	Face Length ≈ Cheekbone Width ≈ Forehead Witch ≈ Jawline && a1 ≈ 90>a2 ≈ a3
Round	(Face Length ≈ Cheekbone Width) > (Forehead Width ≈ Jawline)

III RESULT

The system result is evaluated in two ways. First, the researches evaluate the image face shape using 100 images. Second, the 50 volunteer women and 50 men who were undergraduate students evaluate the system. Prior to the undergraduate evaluation, the researchers explained a guideline how to determine their face shapes. Then the users were asked about their face shapes. The system by answering a questionnaire. The questionnaire was divided into 3 assessments as follows.

A. Researches Evaluation about algorithm

In the system basically the head top point 68 (e.g. Figure 4) different with some factors. For using 100 images we evaluated the point correctness in the system. The accuracy of head point (68) identification algorithm is 75 % of 100 images.

Figure 4. Example on finding all the landmarks detecting

And then researches evaluate the face shape accuracy. The researchers used 100 images which were already know their face shapes. We used 20 images are oval faces, 20 images are round faces, 20 images are heart faces,20 images are oblong faces and 20 images are square faces from the 100 images for evaluating the algorithm of face shape accuracy.

Table 3. Face Shapes Classification

Face shape	Identified face shape
Oval	17
Oblong	15
Square	20
Heart	16
Round	18

The system accuracy of identify oval face shape is 85% , Oblong face is 75% , Square face is 100% , Heart face is 80% and round face is 90%.The result shown in Table 3.

B. System Evolution

System evolution was done by the undergraduate. Prior to the evaluation, the researchers explained a guideline how to determine their face shapes. Then the users were asked about their face shapes. The system by answering a questionnaire. The questionnaire was divided into 3 assessments as follows.

1. User Interaction Assessment System Overview

   The user evaluated the system on (i) an ease of use (ii) a user interaction (iii) an interface design (iv) contents available in the system and (v) a chance of using the system again in the future. The five-level Likert principle to measure the assessment, defining as highest (5), high (4), medium (3), low (2), and lowest (1).

2. Performance Assessment

   To evaluate a performance on (i) a satisfaction of the face shape classification results and (ii) an ability to face shape feature point detection as shown in Fig. 8. Again, the Likert scaling was applied for this assessment.

   The users were asked to determine their face shapes before using the system and the system's prediction matched with the answer given by 40 women out of 50 and 43 men out of 50. It should be noted that the users were not beauty experts and had limited knowledge and experiences to determine face shape.

3. Hairstyle recommendation assessment

   This assessment comprised of a satisfaction of (i) recommended hairstyles and (ii) hairstyle's simulation. Above 5-level Likert scale was used for assessment the recommendation system. According to the result, the system achieved more than four out of five in most of the cases, except the satisfaction of hairstyle's simulation and except an ability to face shape feature point detection.

The evaluation was not implemented in a control room but in different environments and uploaded images get from user devices such as mobile phones, different Cameras and etc. Since the feature point detection was poor this could lead to the poor hairstyle simulation performance.

IV. DISCUSSION

Hairstyle suggestion using image processing is a new area where much research has not been taken. The proposed system has used some new techniques to suggestion Hairstyles and identify face shapes. The system has been tested with a wide variety of different color and resolution images.

This research has done only for color images and aligned images. The failure case is without using aligning images which are difficult to detect faces and give less accuracy for identifying face shapes. Also, the face image's forehead contains without hair, which hiding the research head point feature.

Uploaded images got from users. Those images contain different noise types and lighten, which have given less accuracy of features landmark detection. The images whose values are confined to some specific range of values only. For brighter image will have all pixels confined high values. But a good image will have pixels from all regions of the image. Therefore, we applied histogram equalization improves the contrast of the image. The median filter is used to remove the different types of noises.

Open CV image processing technique selected 60 points in an image of a face that outline the features of the face clearly. Unfortunately, these points did not include the forehead, hence we were not able to identify the top of the face using this algorithm In order to identify a point indicating the hairline. So we have to design an algorithm using Otsu thresholding techniques. This technique does not find points outlining features well, but it can give a good point for separating hair from the forehead, so it was able to find a hairline position.

There were some problems when it comes to identifying the shape of the face. Users are noted beauty experts and they do not have a clear idea about the shape of the face. Even though they found their face shape according to the guidelines given. But it conflicted when it comes to testing using the software. Overall identification of user face shape accuracy of the software was 85%, which percentage proved the accuracy of face shape identification of our system.

Hairstyles gathered from different databases, websites, magazines and beauty experts. The hairstyles categorized for face shapes according to the details of beauty experts, Hong and Derrick.

In the result evaluation phase of the system, the user satisfaction of hairstyle suggestion got 90% percentage. The system can identify face shape more than average 4.00 out of 5.00. It is a good result of this system

In this result evolution, the feature point's detection assessment was at more than average 4.00. The evaluation was not implemented in a control room but in different environments. The feature point detection was poor this could lead to the poor hairstyle simulation performance.

This system can use again the future, according to the evaluation result. Because of that, this system is most important for any generation. They hope to do hairstyles with different hairstyles using that kind of recommendation system.

V. CONCLUTION

The hairstyle recommendation system was presented in this paper. The system classified the user's face shape into five categories which were suitable for hairstyle recommendation. The hairstyle recommendation rules were based on beauty experts' suggestions. Moreover, the system ranked proper hairstyles in the database according to their relevant details. The researched evaluate the face shape accuracy using 100 images. The overall face shape accuracy is more than 80%. It is proving that this system can identify the face shapes very well. The system was evaluated by volunteers on three assessments, including user an interaction, a system performance, and a satisfaction on hairstyle recommendation by the system. The average of three assessments was at 4.05 out 5.00. This is the valuable recommendation system is proven by the assessment of the users. Hairstyles can give value to the human appearance. Many people interact with these kinds of recommendation very easily. This hairstyle recommendation gives various hairstyles to the user and also there have some limitations. Those limitations are can develop with the future works.

ACKNOWLEDGEMENT

There are many people who have directed and have given their helping hand in order to make our research successful. Also, gratefully thanks to Head and all the staff members of the Department of Computing & Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka for the help, guidance and all encouragement through the research period. A special deepest sense of gratitude goes owns of Salons at Negombo for offering knowledge and time to understanding the hairstyles and face shapes.

REFERENCES

Dlib.net. (2018). *dlib C++ Library.* [online] Available at: http://dlib.net/ [Accessed 29 June. 2018].

Hu, L., Ma, C., Luo, L. and Li, H. (2015). *Single-view hair modeling using a hairstyle database*. ACM Transactions on Graphics, 34(4), pp.125:1-125:9.

Visual Studio. (2018). *Visual Studio IDE*. [online] Available at: https://visualstudio.microsoft.com/vs/ [Accessed 29 June. 2018].

Wiesman, L., Booth, J., Manak, C., O'Toole, C., Fitzgerald, B., Romano, E. and O'Toole, C. (2018). *The Most Iconic Men's Hairstyles In History: 1920-1969*. [online] D'MARGE. Available at: https://www.dmarge.com/2016/03/iconic-mens-hairstyles-history-1920-1969.html [Accessed 29 June. 2018].

Sarakon, P., Charoenpong, T. and Charoensiriwath, S. (2014). *Face shape classification from 3D human data by using SVM*. The 7th 2014 Biomedical Engineering International Conference.

Hommel, S. and Handmann, U. (2011). *AAM based continuous facial expression recognition for face image sequences*. 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI).

Xiaoyang Mao, M. T. (2012). *Hairstyle Suggestion Using Statistical Learning.Conference*: Advances in Multimedia Modeling - 18th International.

Sunhem, W., Pasupa, K. and Jansiripitikul, P. (2016). *Hairstyle recommendation system for women*. 2016 Fifth ICT International Student Project Conference (ICT-ISPC).

Xiangxin Zhu and Ramanan,D. (2012).Face detection, *pose estimation, and landmark localization in the wild*, 2012 IEEE Conference on Computer Vision and Pattern Recognition.

Xu, Y., Qiu, J. and Ma, L. (2010). *Measurement and classification of shanghai female face shape based on 3D image feature*. 2010 3rd International Congress on Image and Signal Processing.

Zhiguang Yang, Ming Li and Haizhou Ai (2006). *An Experimental Study on Automatic Face Gender Classification*. 18th International Conference on Pattern Recognition (ICPR'06).

CONSISTENCY IN MULTIPLAYER ONLINE GAME IN CONTINUOUS DOMAIN

HPAI Pathirana[1] and RMCAB Rathnayaka
University of Vocational Technology
[1]pathirana@univotec.ac.lk

Abstract– This paper describes the ways to maintain the consistency in Multiplayer Online Games (MPOGs) in continuous domain. The involvement of computer networking, computer graphics, multimedia, artificial intelligence makes the MPOG environment more complex. As a result, consistency, responsiveness, scalability, fairness and avoiding cheating are examples for available issues in MPOG environment. The consistency is well discussed in the discrete domain, however the attention on consistency of MPOGs in continuous domain is poor and data centric approaches are not adequate for MPOGs due to the game state changes with passing time. Further, the consistency and responsiveness are two important considerations in the MPOG environment, however both cannot be achieved at once. As a result, the different architectures are available to implement game environment, and those are supported by latency handling techniques; time delay, local lag, time wrap, progressive slow down, dead reckoning, gossiping, PREMUB for example. The examples are discussed under each architecture for a better understanding of criticising the architecture. This paper mainly delivers the importance of consistency in MPOG, however it sacrifices the responsiveness in MPOG, as a result both aspects must be considered equally important for Quality of Experience (QoE) of the player.

Keywords- consistency, responsiveness, MPOG, latency compensation, QoE

I. INTRODUCTION

Multi-Player Online Games (MPOGs) are funded more and more by the entertainment industry, and those are getting popular increasingly. The one major reason would be the huge number of players in the internet for entertainment purpose, so it is easy to find a player at any time. The board games, card games, turn based games, word games, role playing games, Role Playing Game (RPG), Real Time Strategy (RTS) games, sports games and First Person Shoot (FPS) games are example for MPOGs, however it is not possible to treat them in same manner due to the different consistency and responsiveness requirements.

Turn based board games, card games are examples for games in discrete domain, since state updates happen as discrete event based on the user initiated operation; no relation of the passing of time. Further, there is no time constrain to present a state of game in discrete domain. These games barely satisfy the requirement of MPOG by maintaining interactive game online, further the inherent causality preservation of games in discrete domain is enough for maintaining consistency in such games. This paper does not focus on such game environments.

In general, MPOG is a type of distributed real time application. RPG, RTS, FPS and sports game are good examples to understand the nature of MPOGs. The document type for interactive MPOG is dynamic, because game state changes for two reasons; relation of the player initiated operation, and the passing of time. The research in consistency maintenance for replicated dynamic document is relatively under explored with reference to the replicated static document.

There are big collection of Massively Multiplayer Online Games (MMPOG) as per the MMPOG Dictionaries online (MMOG Dictionary, 2017; MMORPG, 2017) and those convince the present demand for playing MPOG. For example, the site shows 3686804 members, 974 online users and 952 games on 16th October, 2017 (MMORPG, 2017). MPOG is an opportunity for someone to play game at any time with someone else online. The great level of MPOG collection provides choice for the player.

The MPOG has two platform in general; physical platform and logical platform, whereas the existing physical platform must be supported by the logical platform as the goal of best performance (Hsu, Ling, Li, & Kuo, 2003). In other words, a player represents the physical platform, and a virtual site represents the logical platform. The logical state/virtual site is mainly for consistency maintenance.

The consistency is about the same game state in all instances of the players, whereas responsiveness is concerned on delay for an update to register throughout the network (Gao, Jin, Shen, & Babar, 2017). A shooting dead man is an example for consistency problem; player had shot a man to dead that is responsive enough to reflect locally to determine dead man, however remote player use that dead man to shoot continuously till update is reflected. Highly interactive game environment requires higher responsiveness. Both are important to consider for MPOG environment, however both are are contradictory goals oftenly. As a result, one gets priority over the other depending on the nature of specific MPOG environment. Thresholds for the responsiveness are FPS – 100ms, RPG – 500ms, RTS – 1000ms for example (Gao et al., 2017), more importantly it concludes FPS focus more on consistency over RPG, RTS.

The latency is an inherent major problem to maintain consistency for MPOGs, since multiple players are connected over heterogeneous network; the Internet. Latency is further crucial, as responsive threshold for the MPOG should be within Round Trip Time (RTT) of two game instances (Li, Li, & Lau, 2004). For example, the Final Fantasy XI has discussed in the same paper as one of the popular MPOG in 2004 for example, there is almost one second of delay for having position updates from the other players. Moreover, the restrictions had introduced to reduce those delays to maintain consistency; attack only on objects of enemy, controlled moving speed for objects, but those limits on expected features of the game.

Moreover, the jitter is equally important issue for MPOGs. It happens due to the latency variation. The one common approach on treating jitter in MPOGs is based on maintain threshold for the jitter; farcing to leave the game avoiding counter effects. However it influences on the game fairness and the player QoE negatively (Gao, Shen, & Babar, 2016), and Predictive Modelling of User Behaviour (PREMUB) has been introduced for seamless transmission between player and its intelligent agent as alternative approach. It allows maintaining consistency against poor response.

The rest of this paper further discussed above matters in MPOG environment as follows. The consistency requirements for MPOG environment are disused in detail at next; section 2. Then, the latency handling techniques are discussed as solutions for inconsistencies in two aspects; time based solution, predication based solution in section 3. In section 4, available different architectures are discussed with respect to the consistency concerns with practical example. Finally the paper is concluded with important findings and future directions.

II. CONSISTENCY IN CONTINUOUS DOMAIN

The strictly consistent is about identical replicas at all the time; it is an ideal case. Due to the inherent communication latency among collaborating sites, the strict consistency can never be achieved. There is delay always for propagation from originated game instance to other game instances (Gao et al., 2016; Li et al., 2004; Zhang, Kemme, & Denault, 2008; Zhou, Cai, Lee, & Turner, 2004). Moreover, the human user is able to tolerate a temporary divergence among replicas; as a result the applications do not expect to maintain strict consistency.

For MPOG, due to the dynamic nature of the document, it crucially important to maintain consistency at greatest level; not like discrete domain which focus on correct sequence of independent operation at each site. The consistency of application in continuous domain is focused on execution of operation at correct point in time essentially in addition to the execution in correct sequence as in discrete domain (Mauve, Vogel, Hilt, & Effelsberg, 2004). As a result, discrete domain algorithms to maintain consistency are not sufficient enough for continuous domain to maintain consistency; however those algorithms in discrete domain to maintain consistency are the foundation into continuous domain for introducing relevant algorithms for continuous domain. Further, the continuous domain consistency criteria can thus be regarded as a specialization of the use of the discrete domain consistency criteria.

More interestedly, a single operation is adequately enough for introducing inconsistency for highly interactive MPOG, because there is communication latency in between sites for propagation (Shen & Zhou, 2013). Latency is main hindrance for consistency and responsiveness, and it introduces when the responsiveness threshold of a game cannot be fulfilled within the RTT.

If all the players have received all the operations at any given time, it is ensured consistency with identical state in all the sites at that time (Mauve et al., 2004). However if all the sites has not received all required information at any given time, it is not possible to conclude the state of consistency at given time. Nevertheless, The synchronization of the distinct physical clocks are not able to assist on consistency in continuous domain (Mauve et al., 2004). The same readings of the physical clocks of the all players are important for consistent criteria, if all previous operations have already been received. The same reading of a common wall clock is irrelevant, even though consistency is achieved or not.

The correctness, short-term inconsistency and responsiveness are terms attached with consistency, and those are discussed further in this section at next.

A. Correctness

Beside consistency, the correctness of all the game states is important among all the states of the sites of application in continuous domain (Mauve et al., 2004). The correctness is completely independent of the synchronization of physical clock at distinct sites in theory, however there is no practical issue due to the limitation of human resolution perception as long as the calculation of state is fast sufficiently.

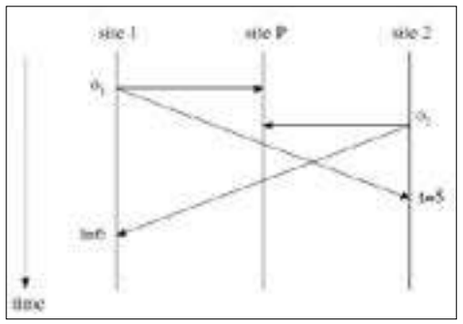


Figure 1. Consistency and Correctness (Mauve et al., 2004)

The consistency and correctness have been described for example as in Figure 1 (Mauve et al., 2004). Site 1 and Site 2 are real to represent a separate instance of continuous distributed application; game instance of player. Site P is virtual site which receives all the user initiated operations; vertical arrow from real site to virtual site describes no delay for updating virtual site. The idea behind the diagram is that the correctness is checked always with the virtual sites which execute operation in order; the correctness has been achieved for identical states in both site instance state and virtual site state. In fact, the consistency is guaranteed when actual sites are identical. As a result, there are situation where consistency is assured, however correctness is not assured. It is important to remember, the correctness is vitally important for MPOG environment on top of the consistency.

B. Short-Term Inconsistency

The short-term inconsistency occurs in a situation where minimum of two sites have different states. It is very common concern, since there is propagation delay from the originated site of the operation to the one other site to appear the operation, so it does not refer the violation of either constancy or correctness under the assumption of primary requirement of application to handle such situation (Mauve et al., 2004). The transmission delay, offset between the physical clock, time gap between operations issued and operation executed are the factors to decide the short-term inconsistency. For example, in Figure 1 the short-term inconsistency appears in site 1; from the time O2 issued to t=6, and site 2; from the time O1 issued to t=5. In the MPOG, this situation is handled by the game software up to some extent, even though it should not be existed in ideal case.

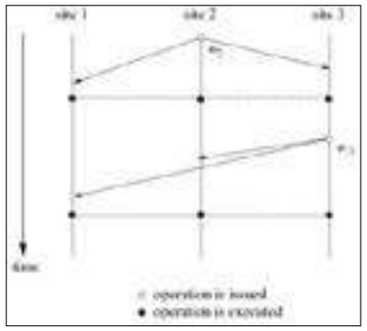


Figure 2. Minimizing Short Term Inconsistencies (Mauve et al., 2004)

The short-term inconsistencies have been minimized in Figure 2 by executing an operation after some time, response time, at the same scheduled time instead of executing at the time of operation issued at the originated site. The response time is estimated assuming the propagation delay of the operation on the other sites, as result it is guaranteed all the participating sites have been received the operation to execute. Further, it is desirable to reduce short-term inconsistency on the cost of responsiveness, while it is not guaranteed on prevention of short-term inconsistency.

i. Game Responsive Threshold

In MPOG, short-term inconsistency can be handled assuring no influence on experience of the player. Due to the different nature of the games, there is specific threshold value to maintain for response time which is game responsive threshold, even though the higher response time can be introduced to have greater emphasis on consistency. There are two groups of MPOGs under this criterion: game either with higher responsive threshold or low responsive threshold.

The higher responsive threshold is applicable for RPG and RTS inclines to have emphasis on consistency, and the lower responsive threshold is applicable FPS and sports game needs greater weight on responsiveness. As a result, short-term inconsistency is not acceptable for games with lower responsive threshold or lower game responsive threshold and there should be other techniques to maintain consistency.

C. Responsiveness

The response time of continues replicated application should be within certain threshold, otherwise local user experiences an unnatural delay in their local site (Mauve et al., 2004). Figure 3 shows site 2 and site 3 issue O2, O3 and those are executed with no response time to assure experience of local user; ideal case. The responsiveness has been optimized in that case at local site, however short-term inconsistency is appeared.

More importantly, the both fidelity criteria; short-term inconsistency and responsiveness, are conflicting goals, because more attention on one criteria influences on the other criteria. In MPOGs, responsiveness is achieved in acceptable level.

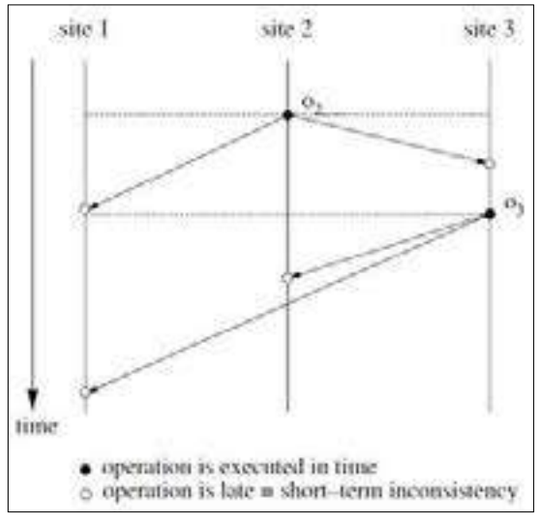


Figure 3. Optimizing Responsiveness (Mauve et al., 2004)

i. Replication

Although the operation is executed without waiting, it is not guaranteed on responsiveness due the inherent communication latency (Mauve, 2000; Mauve et al., 2004). It influences on appearing effect of an operation in one site to appear among other multiple sites for highly interactive MPOG. Nevertheless, there is issue on responsiveness at the local site also with client-server, cloud implementation. The issue at the local site is addressed by maintaining full or partial replica of game state locally, and it is default approach of peer-to-peer approach. It addresses the responsive problem, however it introduces the consistency problem.

D. Causality Preservation

In the continuous domain of MPOG, this introduces interesting problems; due to the short-term inconsistency; if there is no short-term inconsistency, causality is perfectly preserved, since operation executes on all the sites at correct point in time (Mauve, 2000; Mauve et al., 2004). The short delay allows one operation to lead an old concurrent operation. It is essential to delay the new operation until the old concurrent operation to appear and execute. As a result, that new operation introduces short-term inconsistency on future operations in undesirable manner in interactive continuous domain (Mauve et al., 2004). Due to that, the consistency and correctness

criterion are not considered on causality reservation in continuous domain. In the MPOG environment, the causality preservation is not be able to assure due to the above endless delays to assure consistency, as a result short-term inconsistency is not addressed through causality preservation techniques.

III. LATENCY HANDLING TECHNIQUES

Latency is an inherent problem and major hinderance for MPOG environment, because player feel unnatural behavior due to latency (Mauve et al., 2004). The many solutions have been introduced based on assumption on acceptable threshold. Either consistency or responsiveness is focused by the existing solutions. The solutions are based on two aspects; time based approach and predication based approach. Moreover, combination of those techniques is also available. Those multiple solutions are discussed under each category in this section at next.

A. Time Based Solutions

The time based solutions are focused on maintaining consistency, as a result responsiveness impacts negatively.

i. Time Delay / Lockstep

The user commands are not executed immediately, instead those commands are hold by the server for a short period of time, and all the commands are executed in order at the end (Shen, 2017). As shown in Figure 4, client 1 and client 2 send commands to server at different point in time in the timeline, however server holds with the operations of both client to process and execute without introducing inconsistencies and fairness of the game is improved. Unfortunately it introduces matters on responsiveness.

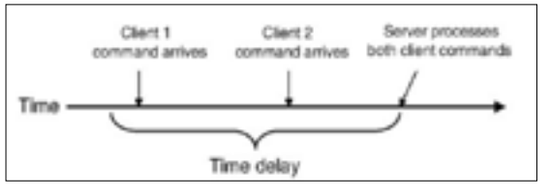


Figure 4. Time Delay (Shen, 2017)

The lockstep is stop-and-wait type approach, which helps on maintaining consistency on the cost of responsiveness(Baughman & Levine, 2001). It is extension for time delay approach with cryptographic approach for avoiding cheating. Further, this approach is not applicable for highly interactive MPOG environment due to the poor responsiveness, even though consistency is maintained.

ii. Local Lag

The local lag has been proposed to address the tradeoff relationship between response time and short-term inconsistencies. The operation is not executed immediately after it is issued by the local user (Mauve, 2000; Mauve et al., 2004), however this concept focus on reducing the responsiveness of the medium. The value for the delay for introduce local-lag must be chosen in the way of neither noticeable nor distraction. Moreover, this approach is not just delaying the operation to execute, but it is all about tradeoff among short-term inconsistency and responsiveness. Finally, it is essential to understand, the local lag sacrifices the responsiveness for better consistency. For example, in Figure 2, the O2 in site 2, and O3 in site 3 have been delayed after they are issued for execution; local lag, to assure consistency.

A minimum and maximum values for the local lag, as per the application tolerance, are determined to finalized the fair value for the local lag for an application based on network delays, jitter, probabilities etc. literature suggest to use 80-100 ms as local-lag with no influence on the user considering even psychological aspects. Finally, the consistency is preserved with fair delay to execute at local site without influencing on responsiveness severely. This approach is not recommended for highly interactive MPOGs due to the poor responsiveness even though consistency is preserved.

iii. Time Wrap

A sufficiently large local lag is able to eliminate short-term inconsistency significantly, but it is not ultimate solution for that due to the jitter, losses over the network (Mauve et al., 2004). As a complement local lag, there should be mechanism to assure consistency and correctness, and time wrap is addressed that problem, however it influence on responsiveness.

Time wrap algorithm waits for predefined period of time to collect local operations and remote operations to store during that short period (Mauve et al., 2004). That short period is based on the requirement of the application based on frequency to calculate new state. In MPOGs, server keeps some snapshots of every recent game states of players with the timestamp (Mauve et al., 2004). If it is necessary to have 25 updates with in a second, the short period of time is calculated as 40 ms. The reconstruction of any inconsistency is conducted with the help of stored interrelated details until state reaches to the current time. The correctness is also assured through complex set of steps with the help of virtual site.

In MPOG, if one player's command involves another player server calculates both player's latency, roll back game state to the previous state of the act moment player sent the command (Mauve et al., 2004). The downside in this approach is the complexity for the computation. The assumption on providing correctness in the time wrap approach reduce the complexity, however using time wrap is a challenge for real time application like highly interactive MPOGs. As a result the complexity of time wrap approach takes into account for giving a solution.

iv. Progressive Slowdown Latency Compensation

This approach addresses both consistency and responsive matters, and this solution only for highly interactive online ball game with two-player (Shen & Zhou, 2013). It includes critical consistency model which expect to have the consistent view at critical state of the game, and it uses this approach for achieving critical consistency. The consistency view is expected have at the critical region as shown in Figure 5, and it is allowed to have inconsistent view at non-critical region. This approach is specific for two player ball games, and each player must have a turn to hit the ball.

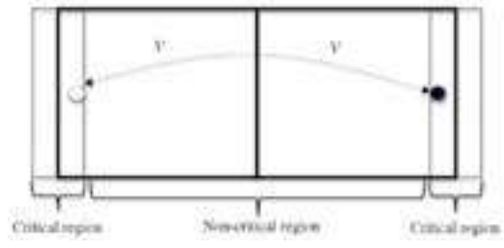


Figure 5. A highly Interactive Two Player Ball Game Court (Shen & Zhou, 2013)

As shown in Figure 6, there are different speeds for the ball as at different location. At the time, ball is close to the player, the speed is maximum for the smoothness and naturalness of the game. The speed is progressively reduced for the inherent latency baring due to the communication for QoE of player, when ball goes away further from the player. The online table tennis game has been use to demonstrate this scenario (Shen & Zhou, 2013). The outcome is satisfied and they are focusing on enhance this for multiple players.

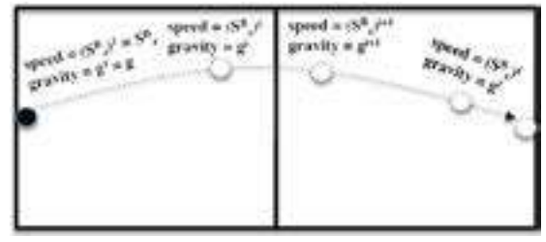


Figure 6. The Progressive Slowdown method for Latency Comparison (Shen & Zhou, 2013)

This method is applicable for peer-to-peer two player online ball games, but not more than two player.

B. Prediction Based Solutions

The state is able to predict, even though there is loose of data. The aim of prediction based approach to reduce responsiveness not giving priority for consistency.

i. Dead Reckoning

Dead reckoning uses to assure, that the consistency criterion is satisfied in distributed virtual environment (DVE) (Mauve, 2000). Both state prediction and state transmission are related with dead reckoning. The behavior of the object in the DVE over the time must be known by the application for this solution. As example, a plane flies in one direction for constant speed or in a given line as calculated considering gravity (Mauve, 2000). The capability to forecast the behavior of an entity is called dead reckoning.

The application maintain single controlling instance at every time for the objects belongs for dead-reckoning (Mauve et al., 2004). The application instances of the pilot are an example from the plane scenario. The controlling application only has capability to change the object state, when the state of object moves over more than a given threshold value

from the predicted state. Then control application sends the affected object related complete state to all other states.

The dead reckoning is able to facilitate consistency, however it is unable to guarantee correctness, because information of the events is not shared with other sites, instead state information is shared (Mauve et al., 2004). Moreover, it is not recommended to send information of the events due to the unreliable transmission. Further this may overload the network if all the sites are going to correct the states, since no reference to follow as correct version; all sites maintain relative state.

The responsiveness is also handled accordingly in dead reckoning. In the game environment, local state is updated with a fixed time delay (100 ms) for providing state update time to reach destination. The choice of time warp for maintaining consistency in the use of dead reckoning has not recommended due to some further implication; responsive might not be able to address.

For the MPOGs, correctness is utterly important to prevent cheating and for fairness. Therefore, it is necessary to address above maters to adopt dead reckoning for MPOGs. Moreover, time wrap would be much appropriate for heavily interactive MPOGs over the state sharing in dead reckoning.

ii. Speculation

This approach predicts future frames which can be possibly appear within RTT (Lee et al., 2015). The state prediction, state approximation, checkpoint and rollback for state, compression state or saving network bandwidth are related concept under this approach to achieve better responsiveness, however the consistency is not addressed.

The Outatime Architecture addresses the latency issue at the server end (Lee et al., 2015) based on speculation concept. Outatime facilitates to hide 120 ms of latency over Internet in the mobile cloud gaming, and it extracts speculative frames to represent potential outcome at next, and transmitting those frames to the other sites before one complete RTT of time (Lee et al., 2015). It helps on recovering wrong/missed speculations. Future state prediction, state approximation with image based rendering and event time shifting, fast state checkpoint and rollback, state compression for bandwidth saving are steps of Outatime. The demonstration had

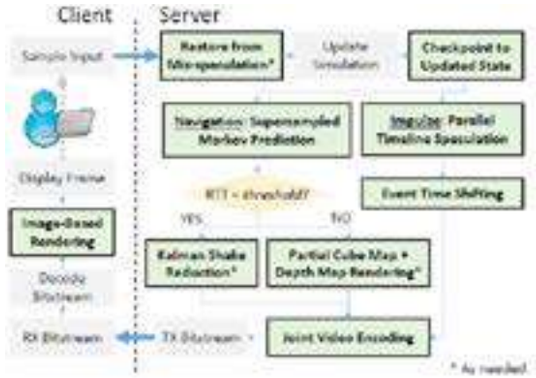


Figure 7. The Outatime Architecture (Lee et al., 2015)

conducted for Outatime on Doom3 FPS game, and it concludes that game providers can implement this for better user experiences.

C. Other Approaches

There are other approaches as combination of both time based and state prediction; gossiping (Lu, Parkin, & Morgan, 2006) and Obsolescence-based Optimistic Synchronization (Shen & Zhou, 2013). Nevertheless, a new approach has introduced with the use of an intelligent agent locally to represent remote user; humanoid bot.

i. Predictive Modelling of User Behaviour (PREMUB)

Intelligent agent is used to model the human behavior in this case, then the humanoid can be used to represent human behavior for any inconsistency (Gao et al., 2017). It improves QoE of the player. The challenge here is that humanoid bot is able to represent the player when there is no inconsistency to imitate the player, however humanoid bot is effective to use against latency hike, known as jitter, instead of leaving the player out.

There is concept to switch for humanoid bot for jitter (Gao et al., 2017). There is no issue when latency is below the threshold of a MPOG, and actual player continue on playing under that situation. There is a background process while continuing normal scenario, each player share its gameplay data with remote site to introduce humanoid bot for them at remote site. At some point in time, latency may exceed the threshold of the game, then play stops communicating with remote site, instead player communicates with humanoid bot of remote site which

is already established locally. It provides time for latency to become value lower than threshold of the game, then players synchronize the states to start playing game, once player starts the game, humanoid bots go into standby mode.

This is a very new approach to improve satisfaction of the player through seamless background operation. In the case study, the pong game has used as MPOG to evaluate this approach, and it is concluded, that there is some promising results on behavioral variables, process, modelling techniques, performance and outcome (Gao et al., 2017), however their verification process is based on historical gameplay data single test subject for short period of time.

IV. MPOG NETWORK ARCHITECTURES

The involvement of interactive computer interfaces, heterogeneous networking infrastructure, hardware support, artificial intelligence and number of players makes the MPOG environment complex. In the literature, the large scope of MPOG provides some guidance for available issues on addressing consistency and responsiveness requirements instead of providing exact solution. The ultimate goal of an architecture is to deliver specific MPOG environment that player is comfortable to use by maintaining consistency and responsiveness; QoE.

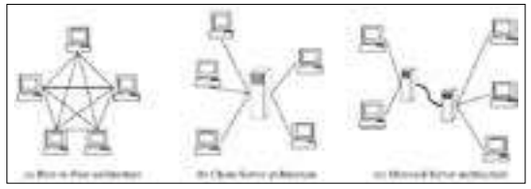


Figure 8. Topology for Multiplayer Online Game System (Hsu et al., 2003)

A. Client-Server Architecture

The Figure 8 – (b) illustrate the topology for this, and the server entity makes decisions. The client-server architecture is better approach for MPOG, since server assist on having consistency among all the playing instances (Hampel, Bopp, & Hinn, 2006). It is based on unicast TCP; a connection oriented lossless transmission is guaranteed. The server maintains the state of the game state applying transformation as required, and each client handles input, output and collision detection for transformation. It favors for consistency ahead of responsiveness; sacrifices the responsiveness to maintain consistency. However client-server communication introduces a latency influencing the user experience due to the poor performance of network communication. Nevertheless, the MPOG industry focus on client-server approach for commercial objectives to facilitate massive number of players at once (Lu et al., 2006).

In general, the client-server architecture is recommended for having data-centric consistency and appropriate for identified MPOG type; no state change against time, acceptable latency in client-server communication (Shen & Zhou, 2013). The client-server is not recommended for more interactive MPOG, because the game state is changed with related to the time progress. For example, P1, P2 are two players of pong game in Figure 9. The local response for P1 has been extended due to the involvement of the server, and the remote response has also been extended compared to peer-to-peer architecture scenario shown in Figure 10. It is very crucial to maintain responsive time within pong game responsive threshold to stay with the game. The server performance is also influence on responsiveness due to the massive number of players.

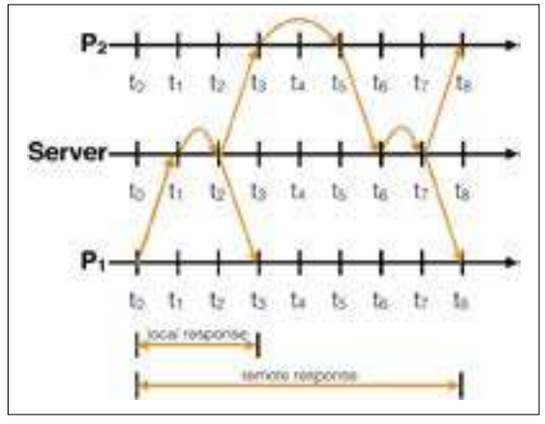


Figure 9. Impact of Latency on Responsiveness (Gao et al., 2017)

i. Distributed Server

The multiple servers are contributed for MPOG, because the single server representation is a bottleneck for MPOG environment (Lu et al., 2006), and the Figure 8 – (c) illustrates mirrored server

architecture with additional server. The influence due to the physical distance of the player is also addressed in multi-server; reduced latency with closest server communication, and server must be connected reliably with a fast connection for better synchronization among them. The responsiveness is enhanced compared to single server environment, and consistency is depend on the consistent communication between servers mainly; if they far away still consistency is issue, even though it is better than having single server.

Further, massive number of players must be supported with cluster of servers utilizing cumulative resources for processing and maintaining (Lu et al., 2006). The clustered-server architecture assists on load balancing. If one server is down due to some reason, there are some other servers nearby to take over the responsivity seamlessly. Clustered server environment is far more supportive for maintaining consistency, whereas responsiveness may also improve depending on the distance between player and server closer the distance is much better.

B. Peer-to-Peer Architecture

The peer-to-peer architecture represents a communication between two players connected somehow. It maintains fully replicated game state, and there is direct communication to maintain consistency of replicas (Diot & Gautier, 1999; Ferretti, 2008). The Figure 8 – (a) illustrate the topology for this, and all the participating entities make decisions. It favors for responsiveness ahead of consistency. As a result, inconsistency appears, and repair inconsistency by using latency-hiding techniques discussed above. So, consistency, responsiveness, scalability, avoiding cheating and fairness are example for challenges in this architecture.

On the other hand, MPOGs in peer-to-peer use Real Time Protocol (RTP) over UDP/IP; data can be loose due to the connectionless unreliable transmission. No central server exists to maintain state of game, as a result each participating entity maintain own game state resulting the greater possibility of inconsistency. It is responsibility of each entity to contribute for synchronization between multiple game states with heterogeneous Internet delays.

The following figure discusses impact of latency using a pong game. There is no any influence on local response since event of P1 reflects locally with no delay, however remote response still has influence, though it is better than client-server architecture. It is crucial to maintain response time within pong game responsive threshold to stay with the game, however that risk is low compared with client-server environment. It further concludes, that it is necessary to maintain different threshold for different architecture in the cost of responsiveness.

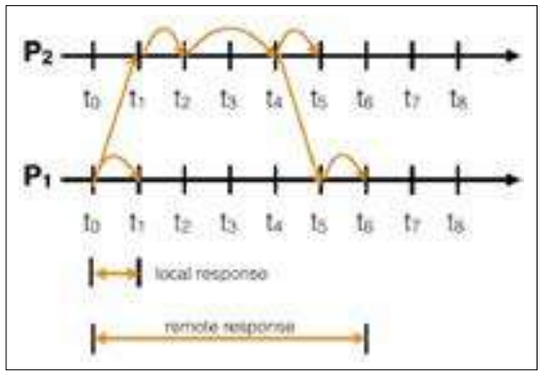


Figure 10. Impact of Latency on Responsiveness (Gao et al., 2017)

Considering the local response time and remote response time among client-server architecture and peer-to-peer architecture, the choice for highly interactive MPOG is peer-to-peer architecture in general, but the consistency can be easily maintain in client-server environment with massive number of player compared to peer-to-peer architecture.

i. Server Mediated Peer-to-Peer Architecture

The disadvantages of peer-to-peer architecture due to the unstructured and decentralized environment; no central coordination, are addressed in this approach (Kwok, Chan, & Cheung, 2005). The server mediated approach has introduced for addressing the consistency issue fair sacrificing on responsiveness. The all nodes in game environment are not players, some nodes represent the role of server as a facilitator/coordinator. In MPOG, even though the server facilities the game environment, that server does not contribute on decision making (Shen, 2017). The inter-trusted peers are grouped around a server for sharing networking and computing capabilities for better experience of the player.

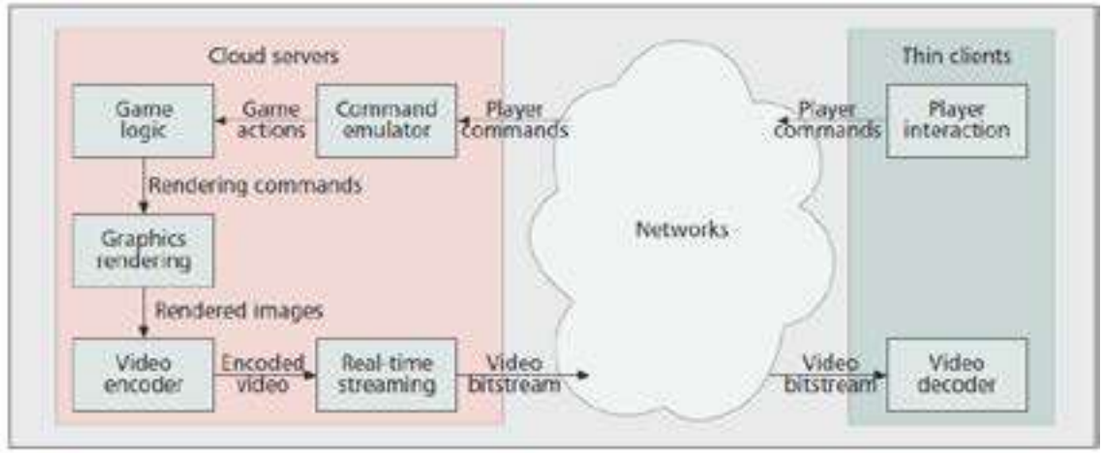


Figure 11: A Cloud Gaming Framework (Chuah et al., 2014)

MiMaze is an example for implementation of such architecture (Diot & Gautier, 1999). It does not maintain a central server to game state, however it uses sever for non-real-time task such as session management for new player. The distributed games must maintain consistency, so it is essential to implement synchronization mechanism at minimum. The acknowledgment based approaches introduces additional communication to the busy Internet (Diot & Gautier, 1999), however dead reckoning algorithm to recover losses and latency of packets has been considered.

C. Cloud Games Architecture

The cloud infrastructure heavily supports for MPOGs on portable devices; smart phones, laptop, tablet (Ferretti & D'Angelo, 2012; Lee et al., 2015). The remote servers is responsible to perform execution of game and rendering as per the input from the client, and thin client displays output frame. Further, it facilitates to play any game at any time ignoring the performance of the thin clients by delegating complex computation into cloud end (Chuah, Yuen, & Cheung, 2014). Further, it helps to save the battery of mobile devices for better user experiences.

Cloud game provides less maintenance, no upfront cost, better scaling (Chuah et al., 2014). Game developers are also further benefited due to the pay as you go concept, since they no need to pay upfront cost to introduce game, however the growing number of players needs more resources which can be allocated instantly. These motivates players to play MPOG and developers to develop games. The MPOG is going to be very big industry.

In Figure 11, a MPOG framework has introduced to illustrate video game environment over the cloud (Chuah et al., 2014), it shows clearly the contribution of cloud server for thin client; all the processing and computations have been delegated into the cloud server.

The massive number of players involvement on MPOG is complex to predict resource requirement to accommodate all the users with better gaming experience, and resource constrains lead for inconsistency, as a solution cloud facilitates scalable resource with no interruption (Chuah et al., 2014). The cloud is very useful to implement the game environment by introducing required physical infrastructure for any architecture delegating complex computation, resource requirement for spike of players, maintaining consistency. Cloud game architecture maintain consistency much better way, and responsiveness depends on the network delay in the Internet access.

V. CONCLUSIONS

The existing literature is reviewed for evaluating consistency of MPOG in continuous domain. . The involvement of computer networking, computer graphics, multimedia, artificial intelligence makes the MPOG environment more complex. But MPOGs are increasingly popular with the interest of the players and the affordable

infrastructures requirement; mobile devices with better performance, countless cloud services, wideband of the Internet facility. However, highly interactive MPOGs have challenges due to the network latency, and it impacts on design and deployment of highly interactive MPOG. The latency handling techniques are used to address issues as shown in Table 1. The requirement of MPOG decides the most relevant approach. Finally latency must satisfy the game threshold to assure QoE of player. Further, different MPOG network architectures are discussed for understanding the support to maintain consistency requirement in MPOG environment.

Table 1. Summary of Latency Handling Techniques

Techniques	Time Manipulation	Prediction Based	Responsiveness	Consistency
Time Delay	x			x
Local Lag	x			x
Time Warp	x			x
Progressive Slowdownx		x	x	
Dead Reckoning	x	x	x	
Speculation	x	x		
PREMUB		x	x	

In summary, the available latency handling techniques and network architecture address consistency requirement of highly interactive MPOGs, however certain solutions only work well with certain architecture as shown in Table 2. The table has been introduced based on present personal understanding on discussed different architectures and latency handling techniques in this paper, however specific considerations should be based on actual MPOG environment requirement instead of architecture to adopt best latency handling technique. Finally, interactive MPOG still suffer due to the consistency issue. It is necessary to conduct comprehensive research focusing highly interactive MPOG approaches.

Table 2. Summary of use of Latency Handling Techniques

MPOG Network Architecture	Time Delay/ Lockstep	Local Lag	Time Warp	Progressive Slowdown	Dead Reckoning	Speculation	REMUB
Client-Server			x		x		
Distributed Server			x		x		
Clustered Server			x		x		
Peer-to-Peer			x	x	x	x	x
Server Mediated P2P			x		x	x	x
Cloud Games					x		

REFERENCES

Baughman, N. E., & Levine, B. N. (2001). *Cheat-proof playout for centralized and distributed online games*. Paper presented at the INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE.

Chuah, S.-P., Yuen, C., & Cheung, N.-M. (2014). Cloud gaming: a green solution to massive multiplayer online games. *IEEE Wireless Communications*, 21(4), 78-87.

Diot, C., & Gautier, L. (1999). A distributed architecture for multiplayer interactive applications on the Internet. *IEEE network, 13(4)*, 6-15.

Ferretti, S. (2008). A synchronization protocol for supporting peer-to-peer multiplayer online games in overlay networks. Paper presented at the Proceedings of the second international conference on Distributed event-based systems, Rome, Italy.

Ferretti, S., & D'Angelo, G. (2012). *Mobile online gaming via resource sharing*. Paper presented at the Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques, Desenzano del Garda, Italy.

Gao, C., Jin, K., Shen, H., & Babar, M. A. (2017). Are you a human or a humanoid: Predictive user modelling through behavioural analysis of online gameplay data. *Advanced Engineering Informatics*.

Gao, C., Shen, H., & Babar, M. A. (2016). *Concealing jitter in Multi-Player Online Games through predictive behaviour modeling*. Paper presented at the Computer Supported Cooperative Work in Design (CSCWD), 2016 IEEE 20th International Conference on.

Hampel, T., Bopp, T., & Hinn, R. (2006). *A peer-to-peer architecture for massive multiplayer online games*. Paper presented at the Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games, Singapore.

Hsu, C.-C. A., Ling, J., Li, Q., & Kuo, C. (2003). *The design of multiplayer online video game systems*. Paper presented at the Proceedings of SPIE.

Kwok, S. H., Chan, K. Y., & Cheung, Y. M. (2005). A server-mediated peer-to-peer system. *SIGecom Exch.*, 5(3), 38-47. doi:10.1145/1120680.1120686

Lee, K., Chu, D., Cuervo, E., Kopf, J., Degtyarev, Y., Grizan, S., . . . Flinn, J. (2015). *Outatime: Using speculation to enable low-latency continuous interaction for mobile cloud gaming*. Paper presented at the Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services.

Li, F. W., Li, L. W., & Lau, R. W. (2004). *Supporting continuous consistency in multiplayer online games*. Paper presented at the Proceedings of the 12th annual ACM international conference on Multimedia.

Lu, F., Parkin, S., & Morgan, G. (2006). *Load balancing for massively multiplayer online games*. Paper presented at the Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games.

Mauve, M. (2000). *Consistency in replicated continuous interactive media*. Paper presented at the Proceedings of the 2000 ACM conference on Computer supported cooperative work.

Mauve, M., Vogel, J., Hilt, V., & Effelsberg, W. (2004). Local-lag and timewarp: providing consistency for replicated continuous applications. *IEEE transactions on Multimedia*, 6(1), 47-57.

MMOG Dictionary. (2017). MMOG Collection. Retrieved from http://mmogdirectory.com

MMORPG. (2017). MMORPG Collection. Retrieved from http://www.mmorpg.com/

Shen, H. (2017). COMP 7722 - *Computer Supported Cooperative Work and Groupware*. Retrieved from Flinders University - Adelate - Asustralia

Shen, H., & Zhou, S. (2013). *Achieving critical consistency through progressive slowdown in highly interactive multi-player online games*. Paper presented at the Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on.

Zhang, K., Kemme, B., & Denault, A. (2008). *Persistence in massively multiplayer online games*. Paper presented at the Proceedings of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, Worcester, Massachusetts.

Zhou, S., Cai, W., Lee, B.-S., & Turner, S. J. (2004). Time-space consistency in large-scale distributed virtual environments. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 14(1), 31-47.

REQUIREMENTS FOR AN ENGLISH-SINHALA SMART BILINGUAL DICTIONARY: A REVIEW

L Samarawickrama[1] and B Hettige
Department of Computer Science, Faculty of Computing, Sir John Kotelawala Defence University, Sri Lanka
[1]*lasikasamarawickrama@gmail.com*

Abstract- This paper presents a review on existing English-Sinhala Dictionaries while considering today's requirements for enhancing English language skills for users who are not fluent in English . Dictionaries are becoming popular as a means for interactive communication between languages. Dictionaries, and language professionals, in general, have long claimed that dictionaries are deficient, especially regarding accessing and updating content. Some authors have also noted that these deficiencies are compounded by the fact that language professionals do not receive proper training in dictionary use, and due to those deficiencies, they do not fully benefit from them. Electronic dictionaries include new search capabilities, not found in traditional dictionaries, that could meet users' needs efficiently. So, this review, it's discusses the features of the dictionaries with a comparison to existing dictionaries. Finally, its state the features that should be added to those existing ones to become a smart bilingual dictionary.

Keywords- Bilingual- Dictionary, Morphology, Context Representation

I. INTRODUCTION

The EF English Proficiency Index("EF English Proficiency Index - Sri Lanka," 2017) ranks countries by the average level of English language skills. According to the EF English Proficiency Index Sri Lanka has a very low proficiency. To enhance English language skills English to Sinhala bilingual dictionary can be used and it significantly effects many activities. In addition to the above, most of the people use a mobile phones and hand-held devices. So, developing an English Sinhala dictionary as a mobile application or that can access through the mobile devices will be worthful.

A dictionary, sometimes known as a guide, is a collection of words in one or more specific languages. ("Dictionary," 2017a). A dictionary can give you a clear definition, grammatical form, comprehensive information on pronunciation, and most importantly carefully chosen examples that show how to use the word. Learners, teachers and textbook designers should be aware of the importance of the dictionary as an important learning device. A dictionary used as a reference, a guide for language teachers and learners with highly confidential help in teaching. It is not for decoding meaning, but encoding is another function of a dictionary.

There are several types of dictionaries. In many languages, words can appear in many different forms, but only the underlined or unconjugated form appears as the headword in most dictionaries. Dictionaries are most commonly found in the form of a book, but some newer dictionaries, like StarDict ("StarDict," 2017) and the New Oxford American Dictionary ("New Oxford American Dictionary," 2017) are dictionary softwares running on PDAs or computers. There are also many online dictionaries accessible via the Internet.

The online dictionaries are limited in some way due to various reasons. Typical limitations are lack of recorded pronunciations (Longman, Oxford, Cambridge), missing phonetic transcriptions (Longman), a smaller number of example sentences. ("Comparative review of dictionaries for English learners | Antimoon," 2017)

Accessing Web dictionaries is lower than a dictionary that's installed on your computer. It's not only about the page loading time and server problems. Often, the interface requires more looking around and clicking. When you look up dozens of words time to time every day , it can really hurt your motivation if your dictionary slows you down. So that it reduces the learning of an language.

Therefore, web dictionaries are a useful resource when you want to get some extra information about a word (e.g. get some more examples, double-check the pronunciation). So with the development of the technology developing a dictionary application which can access through the mobile phone by everyone will be worthful to reduce the distance between languages.

There are a number of online and offline dictionaries available for English-Sinhala language pairs including Malalasekara, Godage, Madura etc. These dictionaries show their features and weakness. Thus, this paper reports an overview of the existing dictionaries including their approaches. In addition, this paper also discusses what are the required technologies and features for a new dictionary that can be used as a tool for enhancing English language skills.

The rest of the paper is organized as follows. Section 2 reports a summary of the historical development of the dictionaries. Section 3 reviews existing dictionaries and their approaches including their features and limitations. Finally, section 4 discusses requirements for a new smart dictionary including new features that can be included.

II. HISTORICAL DEVELOPMENT

History of dictionaries dates back to early centuries. The Akkadian wordlists which was discovered in Ebla (modern Syria) and dated roughly 2300 BCE. A Chinese dictionary in the 3rd century BCE was the earliest surviving monolingual dictionary. The first Sanskrit dictionary, the Amarakośa, written around 4th century CE. Written in verse and it's listed around 10,000 words. The first Japanese dictionary was Niina glossary of Chinese characters. The oldest existing Japanese dictionary, Tenrei Banshō Meigi, was also a glossary of written Chinese. ("Dictionary," 2017b)

A. Dictionaries found in Britain:

The earliest dictionaries in the English language were glossaries of French, Spanish or Latin words along with their definitions written in English. The word "dictionary" was invented by an Englishman called John of Garland , he had written a book Dictionarius to help with Latin "diction". An early non-alphabetical list of 8000 English words was the Elementary, created by Richard Mulcaster. The first purely English alphabetical dictionary was A Table Alphabeticall, written by English schoolteacher Robert Cawdrey.

B. Sinhala Language Dictionaries:

Sinhala is a Sanskrit term that is corresponding to Middle Indo-Aryan word Sīhala. There are four stages in the development of Sinhala language as Sinhalese Prakrit, Proto-Sinhalese, Medieval Sinhalese and Modern Sinhalese. The Sinhalese alphabet, Sinhala Hadiya, is based on the ancient Brahmi script, as are most Indian scripts. The Sinhalese alphabet is closely related to South Indian Grantha alphabet and Khmer alphabet. ("Sinhalese language," 2018)

The Sinhala Dictionary in respect of the Sinhala Language that has evolved over a long period of time through the two thousand years, with various changes and influences. Compilation of Sinhala dictionaries is not an unusual exercise in the Sinhala language. Several dictionaries which are written in the glossary style in Pali and Sanskrit languages are available in the Sinhala literature. Several sinhala dictionaries were written in alphabetical order by scholars such as Kataragama Rajaguru Bandara, Veragama Punchi Bandara, and Ven. Boruggamuve Revata Thera is found in the history of lexicographical works, they are not comprehensive and complete dictionaries. It was with the advent of European nations in the 16th Century, that the need for a modern, scientific dictionary was felt. The initiative was taken by the British. nationals such as Benjamin Clough, Charles Carter and John Callaway in compiling Sinhala - English dictionaries is seen as a result. ("History," n.d.)

III. TECHNIQUES BEHIND THE DICTIONARIES

This section briefly describes some required features and concepts for a Bilingual-Dictionary.

A. Technology

The development of new technologies and the Internet have progressively modified the concept of the dictionary.

Most of the paper dictionaries have been converted to electronic formats and new online dictionaries have been implemented.Based on the technical evaluation, the author distinguishes between online or offline bilingual dictionaries. Offline dictionaries comprise books, pocket electronic dictionaries (PEDs) and PC dictionaries. PC dictionaries include dictionaries in CDROM, floppy disk, and other formats. Based on meta(lexicographic) evaluation,("Lexicography," 2017) this typology distinguishes between electronic dictionaries based on their paper versions, and newly developed electronic dictionaries with different newly added features.As for the online dictionaries, web applications ,mobile applications can be taken.

B. Quality of example sentences

Example sentences are the most important feature of a dictionary. By looking at an example sentences anyone can figure out the answers to three critical questions: 1 What does this word mean? 2 In what situations do native speakers use this word? 3 What words and grammatical patterns typically appear with this word? Examples also give you ready-made phrases that you can select and use in your own sentences.

C. Quality of definitions

In a learner's dictionary,("Merriam-Webster's Learner's Dictionary," n.d.) definitions should tell the learner about the most common situations and contexts in which a word is used.It should give each and every meaning that could be with a word. By having a definition is easier to understand and more useful than a more general, abstract one.

D. Accuracy and completeness of phonetic transcriptions

Dictionaries should list all the common pronunciations of a word in both British English and American English (General American).It will be easier for the user to get through.

E. Quality and coverage of recordings

In most of the online dictionaries, this feature cannot be identified. Audio recordings should sound clear with no technical defects. It should be indicated which transcription is pronounced and the recording should match this transcription. Ideally, there should be a recording for each British and American transcription, as well as for word forms (-s, -ed, etc.)

F. User friendliness

This is an important feature which is necessary for the user.Since every user seeks for an user-friendly interface software should enable the user to perform all the common tasks (searching up a word, playing a recording, etc.) easily and without unnecessary clicking. It should offer all the expected features like selected text input (like copy and paste) and should support all the expected input operations .

G. Search engine

Searching is part of "User friendliness", because technically user seeks for that feature first. But it is so important in a dictionary that it deserves its own category. If a word has many spellings (e.g. hard-pressed/hard pressed), it shouldn't reason out which one you search for.

H. Layout

Definitions, phonetic transcriptions, and example sentences should be laid out in a clean and readable way. The user should not feel overwhelmed by the interface element.

I. Other features

Direction :Direction should be highly concerned whether it's unidirectional, bi-directional or both. Most of the present world dictionaries contain an only unidirectional translation.

Pronunciation: Another feature that will be easier for the user so that user can get the word.

Grammer information: For those who learn English as a language this feature will be more beneficial because grammar is a necessary component in each and every language.

Mode: Most of the dictionaries are Unicode and as an additional content font, images can be used as the mode.

Morphological support: Adding a particular form, shape, structure or an image will be an additional feature which can be included in a smart bilingual dictionary.

Freely availability: Nowadays most of the web and mobile applications are freely available to use and if it's an online dictionary, it will be beneficial if the data usage for one search is less.

IV. EXISTING DICTIONARIES

There are so many existing dictionaries, only a few were discussed below.

Malalasekara dictionary was one of the oldest dictionaries and now in the present, it can be taken as an online dictionary too. (" Malalasekera," 2017). Most of the other dictionaries follow words and meaning available in the Malasekara dictionary to generate their own dictionaries since it contains higher number of definitions.

Derana English to Sinhala Dictionary is a feature-rich dictionary which is developed by the support of the Sinhala community. Derana dictionary consists of more than 300000 definitions. Technical terms from medicine, science, law, engineering, accounts, arts and many other sources can be found in this dictionary. Some additional features of this dictionary are pronunciations, listing most used word at the beginning, and relevant sentences for additional reference (beta). Transliteration facilitates to type Sinhala Unicode in an easy manner. Derana dictionary is the first dictionary in this caliber for the Sinhala language. ("Diana English Sinhala Dictionary," n.d.)

Madura English-Sinhala Dictionary ("Madura English-Sinhala Dictionary," 2017)is a free electronic dictionary service developed by Madura Kulatunga. It is available as computer software, an online website, and an android app. The dictionary contains over 230,000 definitions including various technical terms.

EnSiTip(Wasala & Weerasinghe, 2008) is an English-Sinhala translation tool. The primary purpose of the tool is to assist users who have a basic knowledge of English but are unable to grasp much of what the largely English web has to offer. EnSiTip is an easy-to-use, non-intrusive popup dictionary add-on for Firefox. It mostly lies apparently dormant, until an English lexical item unfamiliar to the user is 'hovered over' in a website.

Bhasha Sinhala Dictionary("Bhasha Dictionary | Sinhala-English Dictionary," n.d.) is Sri Lanka's Official Sinhala-English Dictionary app for Android . It provides you offline dictionary service for English to Sinhala word translations and online dictionary service for both English to Sinhala and Sinhala to English word translations.

Buddi Dictionary("Buddi dictionary," n.d.) is an English to Sinhala Bilingual dictionary developed to provide Sinhala meaning for the given English term, Buddi Dictionary comes with several features. The dictionary operates with three modes namely Unicode mode, Singlish mode, and Image mode. In the Unicode mode,

Figure 1. Sample Output of the Buddi Dictionary

Table 1.Comparision the available features on the commonly selected dictionaries

Dictionary	Malalasek	Madura	Derana	Ensitip	Basha	Godage	Buddi
Technology	Book	Web	Web	Web	Mobile	Book	Web
Direction	E to S	both	E to S	E to S	E to S	E to S	E to S
Pronounciation	No	No	No	No	Yes	No	Yes
Grammer Info	Yes	No	No	No	No	Yes	Yes
Mode	Unicode	Unicode	Unicode	Unicode	Unicode	Unicode	Unicode,Font,Image
Morphological suppo	No	No	No	No	No	No	Yes
Freely Available	No	Yes	Yes	Yes	No	No	No

results provide through the Sinhala Unicode, and Singlish mode provides transliterated text for Sinhala. For that, it uses Sinhala to English transliteration method. In the image mode results are generated through the images. Therefore any simple mobile phone (normal cell phone) can display this information easily. Technically, Buddi dictionary is a kind of word translator than a dictionary. In addition to the Sinhala meaning, it shows conjugation form of the English word. This feature is useful to uses to get more idea on the English word. In addition, this feature helps users to take more grammatical knowledge on that word., To provide these features, the system uses an English morphological analyzing tool that can identify each English word with its grammatical coronary. Figure 1 shows the result of the input word 'books'. According to its simple design data usage is very low than the others. (less than a kilobyte for a single dictionary search)

V. DISCUSSION

This paper reported a brief review of existing dictionaries and available features that can be beneficial for people who are not fluent in English. In addition, this review was efforts of some related research in the use of dictionaries in the context and their existing techniques etc. so, Table 1 compares the available features on the commonly available selected dictionaries. In addition, different findings have been obtained, the first of which is the importance of bilingual dictionaries is to have language translation. When it takes two words it should give the best solution or the meaning, therefore it will be easy for the learner to understand. The variation of word's meaning should be clearly identified. As for an example when it takes the word book it has two meanings so it should be clearly mentioned. The findings give an overall beneficial effect of dictionaries to have and idea about what new features should be added. A great attention can be allotted to dictionary research through means of questionnaires, surveys while accomplishing tasks, to take an idea about the importance of investigating and minimizing the mistakes made. Therefore this behavior will boost up the importance of this tool and encourage to develop warm intimacy with dictionaries and to reduce the language gap. From the review of the literature and the analysis of a set of existing dictionaries, it can be concluded that developing a software-based dictionary will be important. In the present every one uses the mobile phone.so as a further work those online dictionaries can be developed as a mobile application, therefore it will be so much help to get knowledge with regarding difficult words.

As the final conclusion, a simple dictionary incorporating all above mention facilities is required. Especially bidirectional support, selected word support, the pronunciation for the given English word and if the user required more details, then dictionary should be capable to provide sample sentences to get a clear idea about the meaning. As further work of this research, Buddi dictionary can be improved to provide maximum facilities to enhance users English language proficiency.

REFERENCES

Bhasha Dictionary | Sinhala-English Dictionary. (n.d.). Retrieved July 14, 2018, from http://www.bhasha.lk/products/dictionary

Buddi dictionary. (n.d.). Retrieved July 14, 2018, from http://dic.budditha.edu.lk/

Comparative review of dictionaries for English learners | Antimoon. (n.d.). Retrieved October 12, 2017, from http://www.antimoon.com/how/learners-dictionaries-review.htm

Derana English Sinhala Dictionary. (n.d.). Retrieved July 14, 2018, from http://sinhala-english-dictionary.blogspot.com/2017/02/derana-english-sinhala-dictionary.html

Dictionary. (2017a). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Dictionary&oldid=803859323

EF English Proficiency Index - Sri Lanka. (2017, September 22). Retrieved September 22, 2017, from //www.ef.com/epi/regions/asia/sri-lanka/

Gunapala Piyasena Malalasekera. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/ w/index.php?title =Gunapala_Piyasena_ Malalasekera &oldid=787984048

History. (n.d.). Retrieved July 26, 2018, from http://www.dictionary.gov.lk/index.php?option=com_content&view=article&id=2&Itemid=104&lang=en

Introduction. (n.d.). Retrieved July 26, 2018, from http://www.dictionary.gov.lk/index.php?option=com_content&view=article&id=1&Itemid=103&lang=en

Lexicography. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title = Lexicography&oldid=801282848

Madura English-Sinhala Dictionary. (2017). In Wikipedia. Retrieved from https:// en.wikipedia.org/ w / index. php ?title=Madura _English - Sinhala _Dictionary&oldid=788982008

Merriam-Webster's Learner's Dictionary. (n.d.). Retrieved October 12, 2017, from http://learnersdictionary.com/

New Oxford American Dictionary. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=New_Oxford_American_Dictionary&oldid=799828616

Sinhalese language. (2018). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Sinhalese_language&oldid=851575948

StarDict. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=StarDict&oldid=781598448

Wasala, A., & Weerasinghe, R. (2008). EnSiTip: A Tool to Unlock the English Web.

A REVIEW OF MOBILE TECHNOLOGY FOR TEACHING AND LEARNING MATHEMATICS

VO Galahitiyawa[1] and RPS Kathriarchchi
Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka
[1]vogalahitiyawa@gmail.com

Abstract - Today, the claim of Mathematics is more critical than ever. Almost everything in the world is relying on mathematics. The subjects like mathematics can be known as a requisite subject of the school curriculum which is compulsory for students and cannot be ignored. Because, as it is mentioned earlier mathematics can be known as a crucial subject. There is a general opinion that most students do not like mathematics due to many reasons. However, students' dislike for mathematics may damage the collaborative other professions as well. This paper examines the reasons for students' dislike for this subject and will provide the most convenient SMART solution. The methodology used to conduct this research is qualitative and quantitative based survey. The online survey was circulated through Email, Social Media and desk research. The survey was successfully completed by 160 persons securing their secrecy. The shared questionnaire brought strength to the research when identifying the issues. According to the previous work, the technological assistance was the main empirical result that they have suggested in order to minimize this situation. According to the online survey results more than 90% raised their flag on the difficulties that they face during calculations and other algebra by their life time experiences. With the rapid growth of technology and the interest of people towards mobile-based applications, it has been shown that mobile devices can be used as a tool to improve students' academic skills. Therefore, this paper suggests a new approach to the user, through which the user can build up an interest towards mathematics. Also, these can identify the aspects that are involved in teaching and learning of mathematics as well as the possible solutions.

Keywords- Mathematics, Mobile application, SMART technologies

I. INTRODUCTION

Over a decade, the mobile technology, such as smartphones, tablets, different applications, etc. has become a requisite part of the human life in all over the world. With the time to time development of the mobile technology the uses or the services of the mobile device has increased. The mobile technology went beyond the traditional telecommunication. It changed the way of communicating, sharing information and how their works done. Likewise, with the combination of mobile technology and the generation's new cognitive new solutions for the current concept of learning has invented. One prospective solution is Mobile Learning (m-learning).

Mobile educational applications motivate and influence students while they focus on problem solving, memory improvement, reading and writing skills. In addition to traditional classroom contexts, the use of innovative technology in the learning progression and evaluation with m-learning tests improves the time of use according to the needs of the student, by personalizing the experience. M-learning seems gratifying to be able to host features and departments in a wide range of environments that provide a more collaborating learning experience for students. It provides a stress-free environment powered by a particularly designed, user-friendly interface. While it removes the design and layout of the text, hindrance and misperception, information makes it easy to think carefully about certain subject areas like Mathematics (Taleb et al., 2015).

Over many centuries, Mathematics is extended to solve problems. It helps to think logically, to identify a problem clearly and to solve the problem with the appropriate method. Now, much more than arithmetic and geometry, Mathematics has become a necessary discipline which is a powerful tool in the world in order to enhance the perspective of every individual in the world. Rather, mathematics is a study of quality, quantity, structure and the space. Since most of the computational concepts works throughout the mathematical base. With the rapid changes in technology, there will be mathematical changes and logical changes as well. Because of these rapid changes, it would be difficult to anticipate the future skill requirements of people. But, if someone has a proficient base of mathematics, would keep him/her agile enough to face for the rapid changes in technology. Therefore, it would be better to be expertise in mathematics. This review leads to a new approach in leaning mathematics with mobile technology.

II. LITERATURE REVIEW

Mobile learning is one of the emerged solutions to the challenges facing in education sector. Even though the various resources are always available for learning, mobile learning tools offer more options for personalized learning. Mobile learning also can be defined to the use of wireless or mobile devices in order to learn on the move. Mobile learning characteristics in the distance learning contextually has described in a better understandable manner by Park, (2011). Mobile learning in the classroom generally allows students to work on projects, solve problems individually or in groups. There are many formal and informal learning opportunities when you can reach so much of content all the time and anywhere.

After the technological evolution, many technological tools have been presented in tend to increase learning. A study has shown that laptops, mobile tablets, iPod touch and iPads are very widespread devices and announced as the latest technological devices to facilitate mobile learning due to the costs and availability of applications. They are used to collect student feedbacks, refer e-books and websites, record their thoughts, document trips, collect and analyze data, and much more(Mehdipour and Zerehkafi, 2013).

But these devices will be substituted to many other different technological devices in the future with the new emerging technologies. Mobile devices can be used by collaborative teams to switch between relative experience and physical abstraction and high-level ideas and abstraction to switch between relational and sensational events. Because of that the work load is not too high among the teams (Rogers et al., 2010).

Over a decade, ML did not produce a single application for learning technologies, but projected promising circumstances, such as the use of graphing calculators and portable response systems in the classroom, use of PDAs to configure small group work, portable tools for arithmetic skills, portable travel guides and portable tools for basic learning (Balacheff et al., 2009). As it is mentioned, it is a great opportunity to provide a support to a person throughout the technology to their lives, to capture and organize their daily education and lives in order to create and share their wide views to explore surroundings (Sharples et al., 2005)

At most of the higher education institutions which are followed by the elementary schools, mobile learning (ML) is widespread. ML is frequently used to support in the professional industry and the students who follows applied science and formal sciences (Wu et al., 2012). The progression of mobile learning in the widest possible framework and explores the importance of ideas on the current development of mobile learning. The challenge for the mobile learning community is the balance between beholding inner, improving your work, and observing and understanding the framework and importance of this work (Traxler, 2009). The difficulties encountered in assessing mobile learning, defined as a direct consequence of the complex nature of socially perceived mobile learning rather than as a technical phenomenon of people who moves to the advance in daily life by creating a context of self-learning and negotiation through parameters, contacts and technological interactions (Vavoula and Sharples, 2009).

In the modern world of information technology, application of mathematics has become more critical. In that case it is important to improve the overall performance of mathematics among students and undergraduates. Even though it is important, most of the school students and undergraduates are not successful in their mathematical education. The main reason for this situation can be ranged from pedagogically, socially, economically, administrative and the attitude aspect of students and teachers and the strategies used by teachers.

But today, it is in an unsatisfactory level. Because, due to the facility impoverishment, lack of study materials, the

quality of teaching has become bad and low. It is shown below in Figure:1. This issue causes a huge breakdown towards the educational development of the whole world (GEZAHEGN, 2007) . Among the factors, mainly, the low performance in mathematics may causes because of the poor attitudes of both teachers and students towards the subject modules. Both parties cannot underestimate the importance of a good attitude towards mathematics in the learning and the teaching processes. Hence, this is important to commence to study all the factors individually which are influencing the performance in mathematics. Mostly, the scope of higher education is rely on the performance of this subject module(Adino, 2015).

According to the Strengthening of Mathematics and Science in Secondary Education (SMASSE), mathematics and science are the modules that students persistently perform lowly today. Other main reason for low performance in Mathematics is lack of confidence. The students who are not really sure about themselves are feel more likely threatened. Most of the time they avoid seeking help from others as well. Because of that most of them fails in their studies and examinations. The students who got high confidence are more likely to seek help when they need it (MUTAI, 2010).

According to more research findings, there are some more reasons why students get slow passing score in their mathematics examinations. They are mainly. Lack of prior knowledge of students, lack of student labor and their parents' support(Acharya, 2017). When it says prior knowledge, it is the knowledge that the students gained in previous grades. Simply, it can be called as basic foundation in mathematics. If a student erases the knowledge of previous grades, it won't help him to do the present grade. In that case the students' failure rate may increase. Generally, a student's mathematics achievements are depending on the student's laborism. That means a student need to manage some extra time to practice mathematics. Today, most of them are not laborious. Consequently, their failure rate increases. Role of parents is also a main aspect that influences in students' achievements.

(Al-Zoubi and Younes, 2015) also have researched about the low achievements in academics. According to them, the most challengeable situation of both teacher and student, is the low academic achievements in the examinations. In that sense, this problem may derive not only in educational problems but also in social, cultural and psychological dimensions. Guidance or the support is much needed for students to achieve their study goads. It clearly can be seen in the Vygotsky's zone of proximal development.

III. METHODOLOGY

The methodology used to conduct this research is qualitative and quantitative based and the online survey circulated through Email, Social Media and desk research. When it comes to the Mathematical knowledge success factor there are several kinds of questions and problems to be considered. For instance, why people have lacked knowledge of mathematics, why it is so complex to learn and high time consumption of learning due to the complexity.

Today, there is a huge breakdown of learning mathematics among the students. Since they don't see the general requirement of mathematics, they have got a less power of critical thinking and quantitative analytical skills. Not only that, they are lacking with computational skills, problem solving, data analyzing, pattern recognition and learning how to approach and solve complex problems.

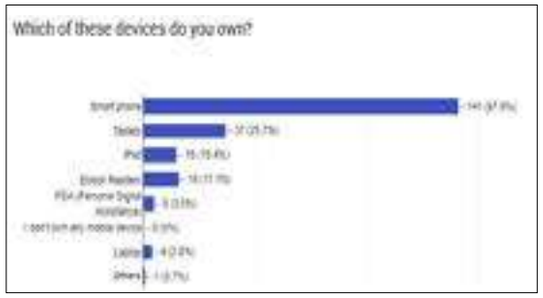
Figer 1. Do you use a mobile device?

Over 160 people participated for this survey via emails, chat rooms and social medias. This research was conducted through an online survey, which included some questions regarding the impact of technology towards the learning factor. This questionnaire was distributed via emails, chatrooms and different social medias. The investigation was successfully completed by over 160 people insuring their privacy. The shared questionnaire brought more strength to the investigation by identifying problems.

IV. RESULTS

Today, mobile device has become an essential thing in life. Out of 161 respondents there were 91 males (56.5%) and 70 females (43.5%). As it is shown in the Figure 1, among those who have submitted their point of view in this survey, there were 46% undergraduates, 22.4% students, 10.6% teachers/lecturers and some other professionals. According to them 98.6% of people owns a mobile device in their day-to-day lives.

As the questionnaires were responded, people use not only the smart phone but they use many other mobile devices in order to ease their basic works. Anyhow the majority with 97.9% are using smart phone. Also, the percentage of people who do not use a mobile device have indicated as 0%. It can be clearly seen in the Figure 2.



Figer 2. The device ownership

As it is questioned in this survey, people have responded that they use their mobile phone or the smart phone almost always (48.6%) and most of them have their phones always with them (60.3%). As well, some people do not hold their phones with them at night and compared to the percentage of having phone always, it is less than 50% (Figure 3).

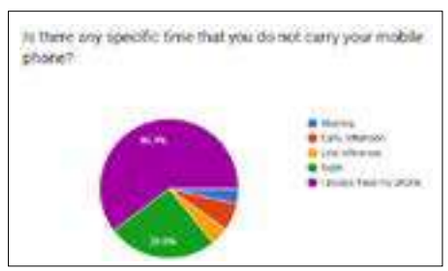


Figure 3. Specific times that people do not carry their mobile phones

Normally, people use their phones in order to communication with others via a phone call. But, today people use their phone most frequently to check out their social media. Some of them use their mobile as a calculator in order to their calculations as well.

When consider about the subject mathematics, it was being a compulsory module that cannot be ignored. Most of the people have completed this module. But their view of this subject is different from one to another. In that case, this investigation has gone through few facts and it was compared to each (Figure 4).

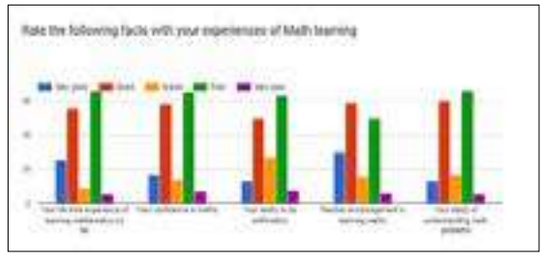


Figure 4. Comparison of peoples' view on Math

Such as;

- The life time experience of learning mathematics so far
- The confidence in Math
- The ability if doing arithmetic
- The encouragement of the teacher
- The ability of understanding math problems

According to the above chart, majority have indicated their knowledge and the attitude towards mathematics as "Poor" while some have indicated as "Good". Also, peoples' the experience of learning mathematics and the understanding the math problem is quite way behind. As for them, they have got a good guidance and a well encouragement when studying mathematics. When considering Figure 3, the overall interest and the attitudes towards this subject is can be seen low.

Not only because the attitude and the interest, but because the indolence of students to refer books have made them to score low mark in examinations. Sometime students have to refer so many books just only to solve one

calculation. As a solution for this issue people have elected to use digital media for referencing over the textbook in referencing. As the Figure 5 shows 87% of respondents have chosen digital media.

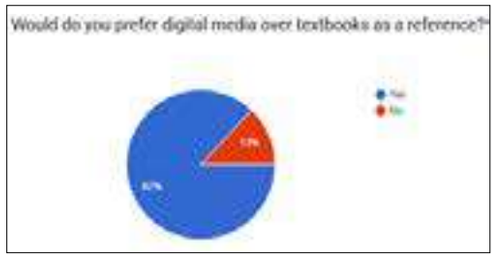


Figure 5. Media Preference

With the development of technology, most of the industries, different sectors have changed. But, the classroom education in Sri Lanka, still in the same level of over a decade. As for the general public who represented the students, teachers and other professional, it is better to use the mobile phone inside the classroom for the learning process as well as for the teaching process. It can be clearly shown in below figure 6.

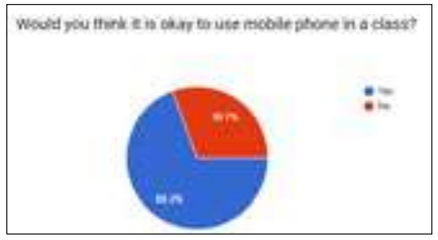


Figure 6. Using Mobile phone in the classroom

In agreement with the previous research authors and as respondence to the survey, the empirical solution that can be seen for the above situation is a technological approach. Because, there are certain amount of people who have been using mobile devices for their learnings and teaching. Correspondingly, it is stated as 92.5 percentage. Also, there were people who suggested that mobile learning should play an important role in Mathematics learning as it is shown in Figure 7.

V. DISCUSSION

Mobile technology is the key point for next generation and let the learning occurs in anytime, anywhere and to

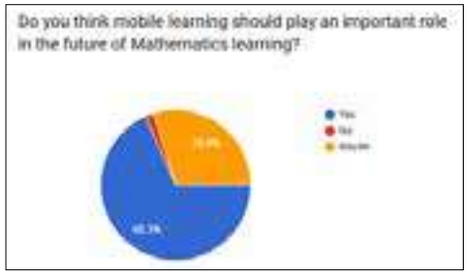


Figure 7. Interest towards mobile learning with mathematics

be an inspiration in a variety of learning circumstances. Mobile learning applications are becoming better known year after year and today they are used by millions of students and educators around the world. Strategies for implementing mobile learning are very important for each nation because of the benefits that this brings. For example, easy access to education and expanded better interaction between instructors and students.

Mobile devices, such as smartphones, personal digital assistants and tablets, could be used to assist students' learning modules, such as mathematics inside of a classroom or beyond the classroom. Inclusion of mobile learning in the educational process, presents some important factors that will examine the implications of using mobile tools and applications in mathematics education at all the levels of education. Mobile learning and related technologies are extensively considered to be increasing ubiquitous in society, especially among young people.

When it comes to be the Mathematical knowledge factor there are several kinds of questions and problems to be considered. For instance, why people have lacked knowledge of mathematics, why it is so complex to learn and high time consumption of learning due to the complexity.

According to the respondents of the survey most of the public individuals have not experienced well while in mathematics education. Most of them are failing in mathematics due to the disability of understanding math problems or the situations. Because of that, students have lost their confidence in learning mathematics. Due to these reasons, there is a huge breakdown of learning mathematics among the students. Since they don't see the general requirement of mathematics, they have got a less power of critical thinking and quantitative analytical skills.

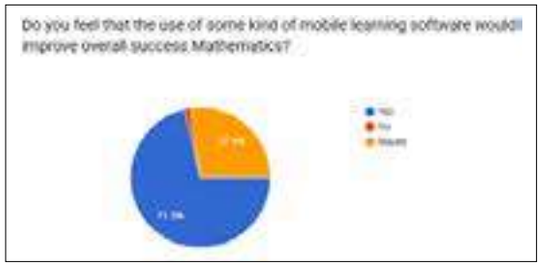
Not only that, they are lacking with computational skills, problem solving, data analyzing, pattern recognition and learning how to approach and solve complex problems.

As a solution for this breakdown in learning mathematics, it can suggest moving forward with the technology. According respondents of the survey which has been taken place on "The usage of mobile devices for learning Mathematics" have marked that having a software tool would be better for math learning. Out of 161 respondents, 112 have marked positive, 43 have marked as may be while only 2 made it negative.

VI. CONCLUSION AND FURTHUR WORKS

With the development of technological change, processes in education, it necessities to be acknowledged that mobile learning should be regarded as a new part of the landscape. Mobile learning offers many opportunities for personal, informal and self-learning. Using mobile devices, students from all ages and backgrounds can work together, interact and learn in different ways.

In addition, the introduction of mobile learning and subsequent technological innovations in schools offers significant opportunities for change and study in teaching practice (Figure 8). Student involvement of trainers will improve learning and teaching opportunities tailored to new trends.



Figer 8. Acceptance of Mobile Learning Software

For a young research area, there is a lot of potential for discrete fields such as mathematics, as well as transforming classes into mobile devices. The result of this study shows that the ability of understanding a problem should be increased as well as the confidence of students. This study was mainly focused on the solution for the breakdown in math learning. According to the previous work done by other research authors and the results of the survey done on this matter shows that adding up mobile technology as a new approach to mathematics learning would be best solution.

The survey results will help to develop the SMART mobile app which can help to solve the mathematical problems such as equation and formulas to be developed in the future as a partial solution to improve the knowledge, confidence and the attitude towards math learning. This concept of mobile learning for mathematics can be expanded further by carrying out this survey for the different schools and the universities. By analyzing results via SPSS statics software can have the best results.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to Mr. RPS Kathriarachchi. I am extremely thankful and indebted to him for sharing expertise, sincere and valuable guidance and the enormous encouragement extended to me. Also, I take this opportunity to express gratitude to all of the KDU IT Department Lectures for their help and support.

REFERENCES

Acharya, B.R., 2017. Factors Affecting Difficulties in Learning Mathematics by Mathematics Learners. Int. J. Elem. Educ. 6, 8. https://doi.org/10.11648/j.ijeedu.20170602.11

Adino, A., 2015. FACTORS INFLUENCING STUDENTS PERFORMANCE IN MATHEMATICS IN KENYA CERTIFICATE OF SECONDARY EDUCATION IN PUBLIC SECONDARY SCHOOLS IN BUTERE SUB COUNTY, KENYA. Partial Fulfillment Requir. Award Degree Master Educ. Curric. Stud. 80.

Al-Zoubi, S.M., Younes, M.A.B., 2015. Low Academic Achievement: Causes and Results. Theory Pract. Lang. Stud. 5, 2262. https://doi.org/10.17507/tpls.0511.09

Balacheff, N., Ludvigsen, S., de Jong, T., Lazonder, A., Barnes, S. (Eds.), 2009. Technology-Enhanced Learning. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-1-4020-9827-7

GEZAHEGN, Y.B., 2007. BARRIERS TO TEACHING AND LEARNING MATHEMATICS IN GRADE FOUR 143.

Mehdipour, Y., Zerehkafi, H., 2013. Mobile Learning for Education: Benefits and Challenges. Int. J. Comput. Eng. Res. Vol. 3, 93–101.

Mutai, J.K., 2010. Attitudes towards Learning and Performance In Mathematics Among Students In Selected Secondary Schools

In Bureti District, Kenya. Partial Fulfilment Award Master Educ. Degree Sch. Educ. Kenyatta Univ. 90.

Park, Y., 2011. A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types. Int. Rev. Res. Open Distrib. Learn. 12, 78. https://doi.org/10.19173/irrodl.v12i2.791

Rogers, Y., Connelly, K., Hazlewood, W., Tedesco, L., 2010. Enhancing learning: a study of how mobile devices can facilitate sensemaking. Pers. Ubiquitous Comput. 14, 111–124. https://doi.org/10.1007/s00779-009-0250-7

Sharples, M., Taylor, J., Vavoula, G., 2005. Towards a Theory of Mobile Learning. MLEARN 9.

Taleb, Z., Ahmadi, A., Musavi, M., 2015. The Effect of M-learning on Mathematics Learning. Procedia - Soc. Behav. Sci. 171, 83–89. https://doi.org/10.1016/j.sbspro.2015.01.092

Traxler, J., 2009. Learning in a Mobile Age. Int. J. Mob. Blended Learn. Vol. 1, 1–12.

Vavoula, G., Sharples, M., 2009. Meeting the Challenges in Evaluating Mobile Learning: A 3-level Evaluation Framework. Int. J. Mob. Blended Learn. Vol. 1, 54–75.

Wu, W.-H., Jim Wu, Y.-C., Chen, C.-Y., Kao, H.-Y., Lin, C.-H., Huang, S.-H., 2012. Review of trends from mobile learning studies: A meta-analysis. Comput. Educ. 59, 817–827. https://doi.org/10.1016/j.compedu.2012.03.016

PERSONALIZED RECOMMENDATION SYSTEM FOR LEISURE TIME ACTIVITY USING SOCIAL MEDIA DATA

M Ramashini[1], DP Jayathunga, A Gowthamy, N Rammiya, and U Kiruthiga
Department of Computer Science and Technology, Uva Wellassa University, Badulla, Sri Lanka
[1]*ramashini@uwu.ac.lk*

Abstract- In today's digital world wherever there's associate degree endless variety of content to be consumed like books, videos, game, movies, music, etc., finding the content of one's interest has become associate degree deadening task. A personalized recommendation system for leisure activities is vital in our social life due to its strength in providing enhanced entertainment. Our system has the ability to recommend leisure time activities, to a new user and others by using their social media data. It gathers all the important information, such as popularity, liking and disliking, required for recommendation. It also takes a minimum of new user information without connecting to social networks. It generates recommendations for the user based on his/her behaviour on social media. Such a system will counsel a group of films, books, music, TV shows, games and places to users supported their interest and private data using Collaborative filtering and Content-based filtering. Similarity, index is measured by using Pearson correlation and Cosine based similarity and Tanimoto Coefficient based Similarity. The planned system has the flexibility to advocate leisure activity to a brand new user furthermore because the others by mistreatment social media knowledge. It effectively reduces the complexity of the search space for users and attracts more and more users to the Internet, which increases the profits of site owners.

Keywords- Recommender system, Cold start user, Big data, Personalized, Hadoop.

I. INTRODUCTION

Recommender Systems ar new generation net tool that gives the power to know a person's style and notice new, fascinating content from the massive variety data} accessible on the online or in alternative electronic information supply supported the pattern between their likes and rating of various things. Similarly, a time off activity recommendation system provides A level of comfort and personalization that helps the user act higher with the system and acquire fascinating suggestions to pay his time off in usefully. This paper presents an outline of the Recommender Systems that ar presently operating within the domain of time off activities like business enterprise, movie, book, music, games, and television programs. we have a tendency to conceive to develop net application and mobile application. With the appearance of massive information, it's become tough to method the huge quantity of knowledge for recommendation (Ankit Kumar Das, 2014).Due to this reason, Hadoop is employed to use for quantifiability, responsibility and quicker process.

The term massive information is outlined by four dimensions pictured by four V's (Volume, Variety, Velocity, and Veracity). Volume is pictured by the number of text information that we have a tendency to ar mistreatment for account to come up with recommendation. selection represents completely different completely different} form of information extracted from different sources like blogs, Facebook, Twitter in addition as totally different review and opinion sites. Speed represents the seed of knowledge generation on the net. these days everyone is connected through the net in addition increasing the recognition of ecommerce sites has become the most reason for increasing the speed of text information generation on the online. Truthfulness represents the trait of the information. Persistently review, opinion, feedbacks ar manipulated or sponsored by totally different stakeholders

of on-line business for his or her personal interest (Holms, 2012) (Sean Owen, 2012) (Jai Prakash Verma, 2015).

Hadoop provides its own classification system known as HDFS (Hadoop Distributed File System). after we deploy our text information on Hadoop classification system, Hadoop distributes all the information in numerous clusters and performs operations parallel. Hadoop conjointly keeps multiple copies of knowledge just in case of hardware failure (Holms, 2012) (Sean Owen, 2012) (Jai Prakash Verma, 2015).

II. PROJECT DEFINITION

Recommendation Systems are primarily directed towards people World Health Organization are having lack of ample personal expertise or ability to judge the possibly overwhelming range of other things that an online web site, for instance, may offer. In most cases, folks are two-faced with decisions and really giant information volumes, and looking out all of them is out of user' capability. This downside is named data overload (A. Felfering, 2007). In the busy world everyone gets leisure time rarely. They wish to spend the valuable free time in useful manner but unfortunately they waste the time and resources by searching better choice and traditional recommendation system suggests items based on rating and reviews. But user may not satisfy with that suggestions.

The recommendation Systems are used in different fields such as tourism, movie, book, music, game, TV shows etc.

Movie Lens is an online movie recommendation system developed by Group lens Research Group, which asks user to rate a few movies on the first login. Through these ratings, the movies are recommended to the user. Collaborative recommendations are provided supported the ratings of comparable users. Grouplens conjointly provides an intensive dataset for movies and ratings in numerous sizes (F.Maxwell harpist, 2016).

MOVREC (Manoj Kumar, 2015) could be a motion picture recommendation system bestowed by D.K. Yadav et al. supported cooperative filtering approach. Cooperative filtering makes use of data provided by user. That data is analysed and a motion picture is usually recommended to the users that square measure organized with the motion picture with highest rating initial. The system additionally encompasses a provision for user to pick attributes on that he needs the motion picture to be suggested.

A hybrid system has been presented by Harpreet Kaur et al. (Harpreet Kaur Virk, 2015). The system uses a mixture of content yet as cooperative filtering formula. The context of the films is additionally thought-about whereas recommending. The user -user relationship yet as user - item relationship plays a significant role within the recommendation.

The Internet is a vast source of information. Whenever any person dreams of planning a tour for himself first thing he is looking for the suggestion from the web and he might faces problems like extensive information and has no idea which information is reliable and which is not. The introduction of digitization in the field of tourism has led to the collection of massive tourism related data. This big data is complicated to process using conventional tools. The data and information are an integral part of any field as it helps in understanding recent trend and patterns. Hence, there is need of a system which will overcome this Problem (Yetis, 2016).

To provide an efficient recommendation to the user, it is of prime importance to understand the user. What are his likings or preferences and much more Social media plays a significant role in understanding user. Social media is a great platform to understand user preferences using which recommendations can be provided. This will ultimately lead to user satisfaction (Miah, 2016) introduces a system which extracts user data from web sources like Flickr and provides recommendation.

There exist variety of book recommendation systems (Greg Linden, 2003) (G. Ramya, n.d.), among that the one utilized by Amazon could be a in style one. Amazon's recommender, as conferred in (Greg Linden, 2003), suggests things to a user that area unit kind of like different users' past purchased and/or rated things, i.e., things that seem within the purchase patterns of assorted users.

(G. Ramya, n.d.)Rely on a ranking-oriented cooperative filtering approach, that considers users' preferences on digital library resources extracted from users' access logs to perform the advice task. This filtering approach overcomes the matter that arises because of the poorness of express users' ratings, i.e., lack of initial info to perform the advice task, in predicting digital library materials of interest to a user.

Park and Yangtze Kiang produce a user-profile P supported individual and cluster behaviour info, like clicks, searching habits, purchases, and interest fields, for generating book recommendations. Using P, the authors calculate the geometrician distance between P and every product profile, that describes product options, and choose merchandise that their geometrician distances ar the nighest to P. extra references on book recommendation systems will be found in Digital Library (Hui Li, 2009). The authors of Ido Guy et al (Ido Guy, 2010) (Maria city Pera, n.d.). Use social-media knowledge to boost the performance of advice systems.

To overcome the issues found in above mentioned existing systems we found out that social network data is extremely helpful for cold start users, users who have not yet interacted with the items. We propose a customized recommendation try and apprehend the characteristics and preferences of the user by collection and analyzing historical behaviour to grasp what reasonably person the user is, what reasonably behaviour preference the user has, what reasonably things the user prefer to share and then on (Mohammad Yahya H. AI-Shamri, 2008) (Xu, 2014) (Jinming, 2010) and eventually perceive that user characteristics and preferences supported the foundations of the platform and suggest the knowledge and product that the user interested (Yan, 2011) (Davidsson C, 2011) it's Associate in Nursing integrated system that may be a combination of a spread of leisure activity recommendations like business, movie, book, music, game and television shows.

In our system a flick recommendation provides A level of comfort and personalization that helps the user move higher with the system and watch movies that cater to his desires. The chief purpose of our system is to advocate movies to its users supported their liked on Facebook and ratings that they supply. In movie recommendation, we wish to use a Hybrid recommender system as it avoids the cold start and sparsity problems inherent with the other models of recommender systems. A hybrid recommender system is one that mixes multiple techniques along to attain some synergism between them.

Book recommendation systems will profit industrial websites, social media sites, and digital libraries, to call a couple of, by assuaging the information acquisition method of users UN agency hunt for books that area unit appealing to them. Our recommendation system that's supported social interactions and private interests to recommend books appealing to users. System depends on the information established on a social networking web site. User primarily {based} cooperative and item based cooperative approaches area unit accustomed suggest the books.

In our system a tourism recommendation is to suggest the destination and places of the user's visit. The Facebook Graph API is used to obtain user information. Facebook is one of the largest social media sites. People are very active on Facebook. This system uses the photos uploaded by the user, the marked photos, and the personal data of the user. Tagged photos and uploaded photos provide details such as the creation time, name, interesting place, the location of the attraction. In this system, the Google Places API is used to get a list of places to visit. After the destination has been selected, the landmark of the destination can be received.

Instead of searching different leisure time activities from many separated system we provide a facility to search different leisure time activities from one system. As a result of this our users can save their time and resources. And users easily get desirable choice among vast amount of suggestions from personalized recommendations in our system. As we said early our system user can spend their leisure time in their preferable way instead of wasting time in searching.

III. OBJECTIVES

A. Main Objectives

To introduce web application and mobile application to recommend the leisure time activities such as tourism, movie, books, music, games and TV shows.

B. Sub Objectives

I. To process information and provide the user with potentially more relevant items by estimating their preferences.

II. To alleviate the matter of data overload, that has created a possible drawback to several web users.

III. To handle the cold start problem effectively

IV. PROJECT DEFINITION

This System facilitate the registered users to login the system through their account and update their user profile. If anyone not register to in system they can register or login through Facebook. At the time system gather and store the information of the user from Facebook and user registration. As soon as user login to the system, leisure time activity categories such as tourism movie book game music and TV show are displayed. User can choose any of them according to their wish.

In tourism recommendation, System captures the current location of the users. System provides two type recommendation.

I. According to the preference.

It recommends most similar users visited places based on the user's preference. Users also can get the detailed information about the places (Distance, Climate details, Description, nearest hospitals, restaurant, police station etc.)

System will tour plan and lists of POI. A purpose of interest or dish could be a feature on a map (or in a very geo dataset) that occupies a selected point, as opposition linear options like roads or areas of land use. The term dish is truly quite inexact, however is widely known by users of satellite navigation systems (SATNAVs), WHO are usually bestowed with choices to indicate or hide points of interest. It is also for geocaching and GIS users, however "POI" takes on totally different meanings in several GIS systems.

II. According to the shortest path.

It recommends the places where old users visited based on shortest distance.

In movie recommendation, for cold start users system will recommend the movies according to the user's preference which are given user during the registration. For registered users system will recommend high rated movie list from similar user preference. Users not only view the movie details but also give the rates to the movie. User can also view the high rate movie list.

In book recommendation system, cold start user get recommendation according to the population estimations and their preference. Ordinary user can view recommendation according to the user based collaborative filter. If user selects a one of the book from the recommendation, he/she can view the available details of the book such as title, author, publisher, rating etc. And also gets book recommendation based on the similarity of the target book.

V. METHODOLOGY

A. Content-based Filtering Systems (CBF based systems): In content-based filtering, items are recommended based on comparisons between item profile and user profile.

B. Collaborative filtering based systems (CF based systems):

Collaborative filtering system recommends items based on similarity measures between users and/or items.

I. User based collaborative filter Recommendations are given to user based on evaluation of items by other users form the same group, with whom he/she shares common preferences (Asanov, n.d.) (Badrul Sarwar, n.d.).

II. Item based collaborative filter

Referring to the fact that the taste of users remains constant or change very slightly similar items build neighbourhoods based on appreciations of users. Afterwards the system generates recommendations with items in the neighbourhoods that a user would prefer. (Asanov, n.d.) (Khyati Aggrawal, 2016)

C. Data Extraction

We collect user details, user likes, movie data and book details from social media data.

D. The cold start problem

The cold start problem appears in two variants. The first variant is related to new users. Whenever a user logs into the system for the first time, the system does not have information about the preferences of that user to provide the recommendations. The second variant is related to new items. When a new item is added to the system, it doesn't have the user ratings. Due to lack of ratings, the item may be placed lower in the recommended items list. To calculate the similarities for neighbourhood, the historic ratings are necessary, and they are unavailable in the cold-start scenario. This makes it difficult to predict ratings for items. [26]

E. Tourism recommendation description

For any person, selecting a destination is a challenging task. The amount of information available on the Internet is enormous which leaves that person confused. In this project, the system takes care of this problem by suggesting an appropriate destination to the individual. It analyses the person's profile and provides personalized recommendations.

F. According to the preference

STEP 1(Profile Analysis): In the first step, new user profile and old user's profiles are compared. The old user list is iterated, and each old user's places visited are compared to new user's visited places. If the place is present in both old and new user's list, this place is added to the common place list. If the place in old user's list is not found in new user's list, the place is added to the different place list.

STEP 2 : After creating a list of common and different places, next step is filter the old user list. Once final old user list is ready, it is sorted according to the maximum number of common places.

STEP 3 : Now, the old user with maximum common places are taken into consideration. The list of different places is taken, and those places are suggested to the new user. STEP 4: Once destination place is finalized, the system will start looking into old user's places. Then it will pick places. To fetch point of interest, the system will use Google places API.

II. According to the shortest path

Owing to the various demerits of pure content-based and pure CF based systems, we have proposed a hybrid recommender system. This hybrid system takes advantage from both the representation of the content as well as the

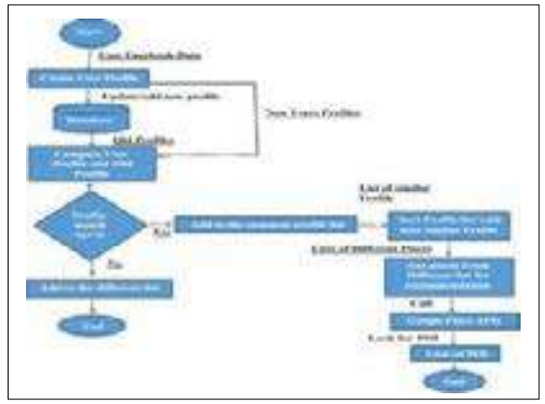


Figure 1. Tourism recommendation overview according to the preference

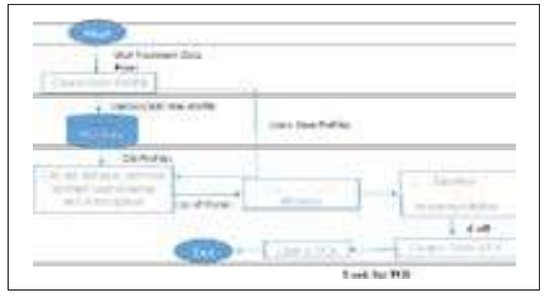


Figure 2. Tourism recommendation overview according to the shortest path

similarities among users. Use a content-based filter to fill the user rating based profile matrix. Get necessary movie content for our dataset. After the preprocessing the movie content database is stored. The dataset consists of a user-rating matrix. Pseudo rating matrix which combines user rating based profile matrix with user-rating matrix. Collaborative filtering is then applied to this full pseudo user-rating matrix to make recommendation for an active user.

Step1: Extract data from Facebook and get minimum information of user's preference about movie. Store the details

Step2: If user is cold start user then create a user rating based profile matrix using content based filtering.

Step3: Weight all users with respect to similarity with the active user. Similarity between users is measured as the Pearson correlation between their ratings vectors. Where, m - total number of items.

$$P_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a)^2} \times \sqrt{\sum_{i=1}^{m}(r_{u,i} - \bar{r}_u)^2}}$$

Step4: Select n users that have the highest similarity with the active user.

These users form the neighbourhood.

Step5: Compute a prediction rate of each movie for current user from a weighted combination of the selected neighbour's ratings.

Predictions are computed as the weighted averages of deviations from the neighbour's mean:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{i=1}^{m}(r_{u,i} - \bar{r}_u) \times P_{a,u}}{\sum_{i=1}^{m} P_{a,u}}$$

Where,

n - Number of users in the neighbourhood.

Step6 : Sort the movie list according to the rate prediction. Step7: Display the top rate movie; there is an option user can see the top rate movie.

Step7: If user is old user then load his/her pseudo rating matrix; go to Step: 3

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}}$$

Step8: In Step7 user click that movie, Display the similar movie using content based filtering based on rate. Step9: Finish.

H. Book recommendation description

Book recommendation methodology is created newly by combination of two existing methodology (Ankit Kumar Das, 2014) (Ishwari Kulkarni, 2017) and some features are added. Book recommendation relies on Item-based collaborative filter approach and user-based collaborative filter approach. Cosine based Similarity, Pearson Correlation-based Similarity and Tanimoto Coefficient based Similarity methods are used to find the average similarity.

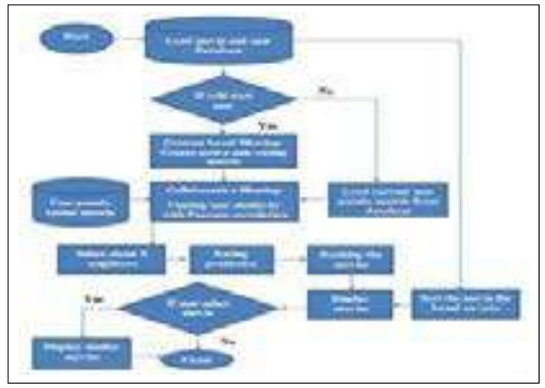


Figure 3. Movie recommendation overview

i. Popularity Estimation

The third type of recommendation algorithm is based on popularity. In general, items that most people are interested are usually a safe choice when you don't know her particular preferences. In the case of new user, popularity becomes one possible solution to the cold start problem. Popularity of the book can be defined as the number of users that have rated that book (Ishwari Kulkarni, 2017).

ii. User type Classification

We classify the user as new user and ordinary user. If user is new user then popularity estimation algorithm is applied as collaborative filtering is not suitable. If user is ordinary user then both popularity estimation and collaborative is applied. If user has not rated any book then it is considered as new user otherwise it is ordinary user (Ishwari Kulkarni, 2017).

iii. Pearson Correlation-based Similarity

The Pearson correlation of two series is the ratio of their covariance to the product of their variances. (Ankit Kumar Das, 2014).

In this case, similarity between two items I and j is measured by computing the Pearson-correlation I, j. Let the set of users who both rated I and j are denoted by U then the correlation similarity is given by

iv. Cosine based Similarity

In this case, two items are thought of as two vectors in the m dimensional user-space. The similarity between them is measured by computing the cosine of the angle between these two vectors. Formally, in the m×n ratings matrix, similarity between items I and j, denoted sin (i, j) is given by where — • || denotes the dot-product of the two vectors (Ankit Kumar Das, 2014).

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

v. Tanimoto Coefficient based Similarity

Tanimoto Coefficient Similarity is an implementation which does not take into account the preference values specified for the users. It is based on (surprise) the Tanimoto coefficient (Ankit Kumar Das, 2014).

Similarity = 2 • similarity – 1

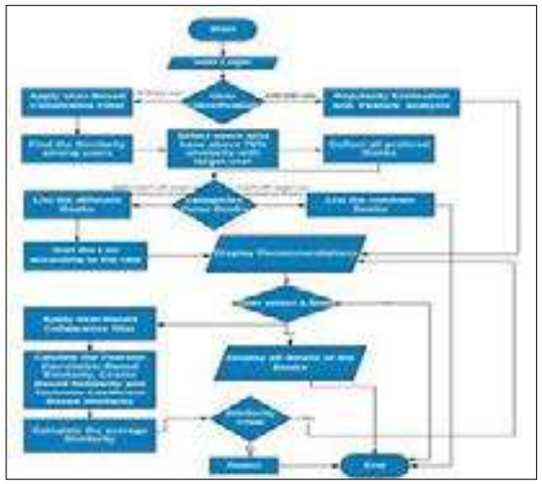


Figure 4. Book recommendation overview

Step1 : Login user classify based on user type classification method. If the target user is a new user follow Step: 02. If user is an ordinary one follow step: 03.

Step2 : Books are recommended to the new user according to the popularity estimation and features that are given by user.

Step3 : For ordinary user, Apply the User-Based Collaborative filter approach and find the similarity among users. Select the users who have similarity above 75%. Collect all books which are preferred by the selected users. From These Books, Separate the Books that are not match with target preferred Books.

Step4 : Separated Books are sorted by rating. Books are recommended by this sorted list. If user select one book from them then Display the available details of the books and go to step: 05 else go to step: 06

Step5 : Apply the Content-based –Filter approach then calculate the Pearson Correlation-based Similarity, Cosine based Similarity, and Tanimoto Coefficient based Similarity individually. And find the average similarity. If average similarity is above 75% then that books are recommended and go to Step: 06 Else go to Step: 06

Step6 : End the process.

I. Methodology for overall Recommendation System

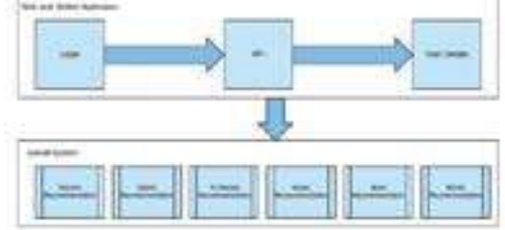


Figure 5. Methodology for overall recommendation system

VI. CONCLUSION

This paper explains how quality recommendations can be given to the user by taking minimum input from users. The system consists of Collaborative Filtering, Content Based Recommendations and Hybrid Recommendations which provide recommendations to the user. Hybrid Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional Collaborative and Content Based Recommendations. To study the user preference we have to use social media data. Our system is a best choice for users who prefer different leisure time activities.

ACKNOWLEDGEMENT

We have taken efforts in developing this project. However, it would not have been possible without the kind support and help of many individuals, faculty of science and technology and department of computer science and technology. We would like to extend our sincere thanks to all of them.

We are highly indebted to our supervisors Ms. Poornima Jayathunga and Ms.Ramashini Murugiah for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project. We would like to express our special gratitude and thanks to our dean of faculty of science and technology and head of department of computer science and technology for giving such opportunity.

Our thanks and appreciations also go to our university in developing the project and people who have willingly helped us out with their abilities.

REFERENCES

A. Felfering, G. F. a. S. T., 2007. Recommender systems. In: s.l.:IEEE Intelligent systems, pp. 18-21.

Ankit Kumar Das, P. K. T., 2014. *Book Recommender System Using Hadoop*, New Delhi: Department of Computer Science and Engineering,Amity school of engineering and technology.

Asanov, D., n.d. *Algorithms and Methods in Recommender Systems*, German: Berlin Institute of TechnologyBerlin.

Badrul Sarwar, G. K. J. K. a. J. R., n.d. *Item-Based Collaborative Filtering Recommendation Algorithms*, Minneapolis: GroupLens Research Group/Army HPC Research Center.

Davidsson C, M. S., 2011. Utilizing implicit feedback and context to recommend mobile applications from first use.

F.Maxwell Harper, J. A., 2016. The Movie Lens Datasets. 5(4).

G. Ramya, K. R., n.d. Personalized QOS based Ranking Approach for Cloud Service Selection. *International Journal of Computer Applications*, October, Volume 127, pp. 0975-8887.

Greg Linden, B. S. a. J. Y., 2003. *Amazon.com Recommendations*, s.l.: IEEE Computer Society.

Harpreet Kaur Virk, E. M. S., 2015. Analysis and Design of Hybrid Online Movie. April .5(2).

Holms, A., 2012. Hadoop in Practice, s.l.: Manning Publications co.

Hui Li, Y. G. S. K., 2009. *A Review of Digital Library Book Recommendation Models*. [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1513415 [Accessed 15 December 2009].

Ido Guy, N. Z. I. R. D. C. E. U., 2010. *Social Media Recommendation based on People and Tags*, Haifa: IBM Research Lab.

Ishwari Kulkarni, P. G. K., 2017. Book Recommendation System Using Apache Spark. *Industrial Journal of Innovative Research in Computer and Communication Engineering*, April , 5(4), pp. 2320-9801.

Jai Prakash Verma, B. P. P., 2015. *Recommendation System with Hadoop Framework*, s.l.: ResearchGate.

Jinming, H., 2010. *Application and research of collaborative filtering in e-commerce recommendation system*, s.l.: s.n.

Khyati Aggrawal, Y. S. G. J., 2016. Movie Recommandation Using Hybrid Recommendation. *International Jourrnal on Recent and Innovation Trends in Computing and Communication*, 4(12), pp. 206-209.

Manoj Kumar, D. A. S. V. K. G., 2015. A Movie Recommender System. August.Volume 124.

Maria Soledad Pera, Y.-K. N., n.d. *Generating Personalized Book Recommendations Using Data Extracted from a Social Website*, Utah: Brigham Young University.

Miah, S. J. V. H. Q. G. J. a. M. M., 2016. *A big data analytics method for tourist behaviour analysis*, s.l.: Information Management.

Sean Owen, R. A. D. F., 2012. *Mahout in Action*, s.l.: Manning Publications co.

Xu, Q. W. J. C. Q., 2014. *A novel mobile personalized recommended method based on money flow model for stock exchange*, s.l.: s.n.

Yan, B. G. C., 2011. *AppJoy :personalized mobile application discovery*. s.l., 9th international conference on Mobile systems, applications, and services.

DATA HANDLING AND MAINTAINING DATA CONSISTENCY IN SCALABLE REPLICATED MICRO-SERVICES

WMNKGTL Weerakoon[1] and Banage TGS Kumara
Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka
[1]*thamira1005@gmail.com*

Abstract- Monolithic Architecture helps to develop server-side enterprise applications. But, it views as a "big ball of mud". That indicates monolithic architecture has many drawbacks. Introduction of cloud based microservice architecture can solve these kind of drawbacks. Microservice architecture helps to scale an application. Most of the applications write less data than reading of that data. Scaling of read model and write model separately is very important. But, scaling applications using microservice architecture is very hard. Further, applications cannot simply use a local ACID (Atomicity, Consistency, Isolation, Durability) transaction. Read part is scaling to more replicas. Thus, maintenance of the data of all replicas in same level is important. Replication of read model and maintenance of data consistency which would provide experimental insight still needs to be developed. To bridge this gap, development of an architecture based on messaging with RabbitMq publish/subscribe method, event sourcing and Command Query Responsibility Segregation (CQRS) with axon framework is used in this study. Evaluation of this architecture was done by replication of the read model using Docker and Docker-compose. Further, we have analysed data consistency in our experiments.

Keywords- Monolithic Architecture; Microservice architecture; Event sourcing;

I. INTRODUCTION

Monolithic Architecture helps to develop server-side enterprise applications. The major advantage of the monolithic architecture is that most applications typically have a large number of cross-cutting concerns, but in Monolithic applications all components are bundled together as a single application, which most people now have come to view as a "big ball of mud". That says monolithic architecture has many drawbacks (Richardson, 2017). Developers move to microservice architecture to solve these drawbacks.

Substantial development of Service Oriented Architecture leads to the rising of Microservice architecture and this focuses on specific aspects, such as decomposition to small services, scaling of the application, improve agile practices for development, independent deployment and separate testing methodologies, continuous delivery features and usage of infrastructure automation, decentralized data management and decentralized

governance among services. Existing researches investigate the characteristics of Microservice architectures. They defined the main facts that need to design Microservice architectures (Alshuqayran, et al., 2016), (Di Francesco, et al., 2017), (Pahl & Jamshidi, 2016).

Data consistency is the most critical part of the microservice architecture. Therefore, it is important to maintain the data consistency. Creation of data in an application happens less than reading of that data. For an example, let us consider a scenario of reserving an airline ticket. Here, writing operation is the reservation of the Seat. Only one person can do this reservation at a time of selected seat. However, many people access the system to check the availability of that selected seat.

This checking of the availability is the reading operation. Therefore, developer needs to pay particular attention to reading performance. However, writing and reading performances cannot be independently separated. Here, we can use CQRS pattern.

The CQRS principle advises separating writing operations from reading operations effectively. Writing part is identified as 'Command-Model' and reading part is identified as 'Query-Model' (Fowler, 2011).

Further, Command-Model must atomically publish events to the all Query-Model whenever their state changes. Event sourcing persists the state of a business entity. A new event will be added to the list of events with the change of state of business entity. Saving an event is just a single operation and especially atomic. The application has the ability to reconstruct the current state of an entity by replaying the events. In this paper, we propose an approach to handle data and maintain data consistency in scalable microservices using CQRS and event Sourcing. The rest of this paper is organized as follows. In section II, we present the problem and motivation. Section III discusses preliminaries. Section IV Methodology. Section V discusses our experiment and its evaluation. Finally, section VII concludes the paper

II. BACKGROUND AND MOTIVATION

Some business transactions span over multiple services. Thus, we need a mechanism to ensure data consistency across the services. For an example, let us consider the previous scenario of reserving an airline ticket. To reserve a ticket, a customer should have at least the credit amount

equals to air ticket price and service charges of the credit card. If the system approves the transaction without having the required credit amount by mistake, then it will be a critical error, which will violate the reliability of airline service. Since ticket, reserving data and customer's credit amount are in different databases the application cannot simply use a local ACID transaction. CQRS can be used to achieve ACID transaction in distributed microservice environment (Melnik, et al., 2012).

According to microservice architecture database per each Microservice is very important. Thus, if we replicate Microservices, every Microservice needs its own database. Assume we have an application with replicas and this application requires new data, which should be added through HTTP request. Once we add the new data to one replica, it needs to synchronize with remaining replicas Fig. 1 illustrates a scenario where synchronization of data is still in progress. In this instance, if someone tried to retrieve data from the third replica, then data retrieved will not be up to date as third replica is still in synchronization process. In that case, if doesn't have eventual consistency. Thus, this research is done to propose a new architecture to solve this problem.

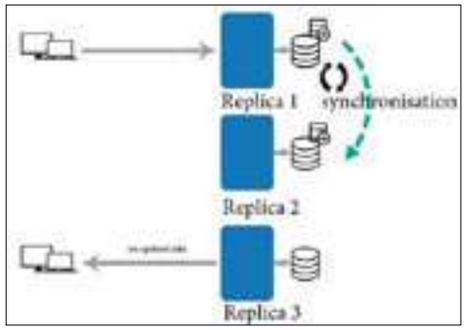


Figure 1. Command Query Responsible Segregation pattern

III. PRELIMINARIES

A. Command Query Responsible Segregation

CQRS is a pattern that use a different model to update information and read information. For some complex operation handling situations, this separation can be valuable. This helps to allocate resources independently to fulfil on-demand provisioning of computing resources. Figure 3 displays the CQRS pattern (Data read part and write part are separately denoted).

Fig 3. shows that application is divided to two parts. One for data reading and one for data writing to database.

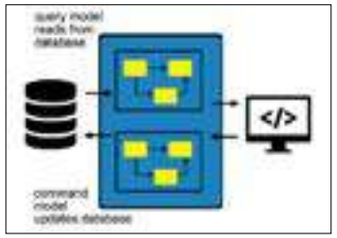


Figure 2. Database level Synchronisation

According to CQRS pattern data read part is called as query model and data write part is called as command model.

B. Command model (Write model)

Commands are responsible for introducing all changes to the system. They lead to change the state of the system. It indicates requests to domain. While request indicates command, it may be accepted or rejected. An accepted event leads to change in the database. Reject command indicate an exception and rollback the system to stable state. Command should not return any value. If there is a distributed system, command emit zero or more events being emitted to incorporate new facts into the system.

Through replicating the instances of read models and by balancing the load, scalable processing could be achieved. Each node could manage its own databases instance with the help of a complete model. Command need to update all the replicas at the same time to manage data consistency.

According to seat reservation system, following two commands help to state change.

- SeatReseveCreateCommand
- SeatReservationUpdateCommand

C. Query model (Read model)

Query is a READ operation. It responsible for reading the state of the system, analyse aggregates, get appropriate data to query. As the response, data is transformed to most useful format. It can be JSON/XML or HTML. According to CQRS theory, query model cannot make any change to the database. In order to achieve the scalability, replicate the read model and introduce database to each replica. It leads to manage request load. Requests can be distributed as a round robin method to all replicas. For some efficient read manner, we can use different kind of databases to this read model. For example, Redis is more efficient than MySQL in big data analysis because more speed is required to analyse more data. Some query models can be implemented with different databases according to requirements.

D. Events

Events are treated as notifications. If something happens, event is responsible to notify that to other interested parties. Command emit 'event' if there is any state change. If all the events are logged in a database, that events can be replayed and check all the state of change. According to seat reservation system two events emitted by commands are;

- SeatReservationCreateEvent
- SeatReservationUpdateEvent

E. Event Sourcing

Event Sourcing is a major part that helps to improve reliability and consistency when CQRS is considered. Event sourcing is a simple theory that logs the state changings. According to seat reservation system events are produced, that represent every change made in the system. Every change is logged according to event sourcing theory. If events are replayed from beginning to current stage that stored in the event store, it helps to understand, how system has changed and will get to the current state of the system. It is similar to technology used in version control systems such as GitHub, BitBucket. In version control systems commit log is used to understand how the code changed by contributors. If there is any conflict, user can replay and can get a stable state. A business object is persisted by storing a sequence of state changing events. Whenever an object's state changes, a new event is appended to the sequence of events. Since that is considered as a one operation, it is inherently atomic. An entity's current state is reconstructed by replaying its events.

F. Eventual consistency

When an application makes a change to a data item on one replica, that change has to be conveyed to the other replicas. Since the change conveyed does not take place at once, there is a time period in which some replicas have the most recent change while other replicas does not. At this time period replica will be mutually inconsistent. However, at the end the changes will be conveyed to all the replicas, and hence the term "eventual consistency". Following Fig 3. shows how eventual consistency is achieved in current study.

According to Fig 3. command handler distributes all new reservation request to the replicas in same time. Steaming and queue ensure all the request are distributing to the replicas.

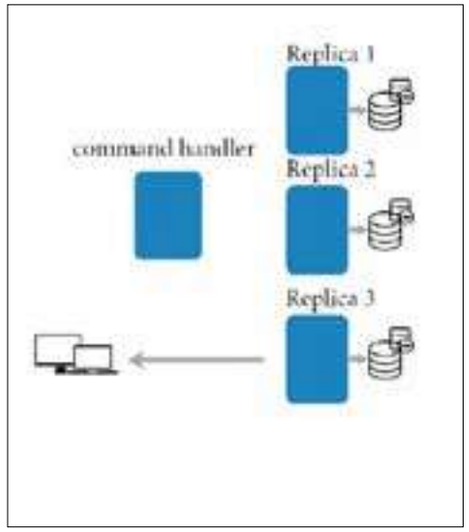


Figure 3. Data distribute to all query model replicas by command model

IV. METHODOLOGY

The purpose of this research is to describe effective and efficient way of "Data Handling and Maintaining Data Consistency in Scalable Replicate Microservices". For that, we applied the Microservice architectures for the proposed system. Then the system is tested for different architectures and most efficient architecture is selected.

In this research, seat reservation system that is applicable to places such as airport or cinema was first selected. This scenario represents a situation where more people access the system at the same time to check the availability of the seat. Some of the requests come to server to only check the availability but the possibility of reserving a seat is less. From that request, few of them reserve the seat. The people who visit the system to check the availability introduce more traffic to the system. An application was developed using CQRS pattern to meet with the requirement of above scenario.

To develop this main application CQRS pattern and Event Sourcing Methods were used. Following technologies were used in order to implement the architecture; Spring-boot 2, MySQL, Axon framework, RabbitMq, Docker and Docker-compose. Fig 4. shows the basic structure of the proposed architecture. Here, seat booking was selected as the core service which acts as the Command-Model. Seat Reserving acts as replicated Query-Model.

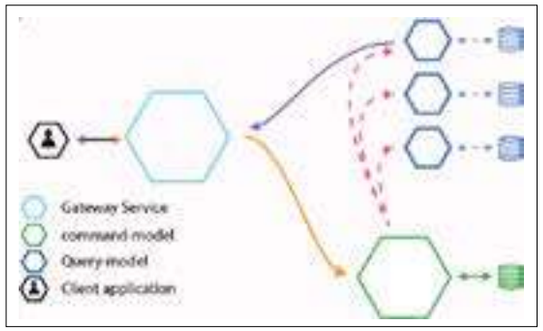


Figure 4. Basic Architecture

Microservice gateway pattern was the main architecture used to implement this system. Gateway service helped to select one instance of seat reservation service, that provides available seats. If a seat is booked from one service, all services need to be updated at the same time by publishing to multiple subscriber pattern. To accomplish this, two 'distribute messages pattern' were used. Those patterns were simple HTTP request and Message Streaming. The results of these two patterns were then compared. After comparison of results, problems related to that scenario were identified. The best pattern was selected and the efficiency of the pattern was maximized.

A. HTTP Request

Implementation of application was done using "Spring boot" web framework and "Rest Template" to test efficiency of simple HTTP messaging. As the first step of implementation, Spring RestController class was created to catch a message. The application then distributes the message to Query-Model. Query part catch the message and save it in the database. A loop is used to send thousands of requests to query module. The time of massage sent (t1) and saved (t2) should be recorded in the database. Then the latency (L) is calculated using following

(1) (Latency is the time taken for massage sending).

L = t2 - t1 (1)

Graph between L and total time taken for send 1000 message is plotted.

B. Message Streaming

The procedure done using HTTP request can be reproduced using data streaming method. This carried out using spring data streaming with RabbitMq message broker. Results showed that data streaming method is the most suitable method for efficient and 'eventual consistent' data transaction. Thus, Message Streaming method is selected to implement CQRS and event sourcing.

V. RESULTS AND DISCUSSION

When implementing the Seat reservation system based on the API gateway pattern, we have to consider about the communication between microservices. When Scaling of Query model and command model is done, those models need to communicate to share the data using HTTP or message stream. Selecting the best communication method is the main part of this research. Fig 5. shows the Latency of two methods used to send 1000 messages using loop. Latency of Data streaming method is illustrated in yellow (Series2) colour line. Orange (Series1) colour line illustrates the latency of HTTP messaging. According to the results, HTTP shows comparatively low latency, so that method is efficient than Data streaming method.

Looping time in HTTP method is higher than the streaming method. According to the results looping time in stream is 707ms and looping time in HTTP is 8491ms. Hence, we consider about looping time stream method is better than HTTP method.

Delay time in HTTP method is lower than the streaming method. According to the results, delay time of streaming is 3266ms and maximum delay time of HTTP is 33ms. When we consider about delay, HTTP is better than streaming.

The difference between HTTP and streaming method is HTTP method waits for the response of recently sent message to send the next message and this is how HTTP method ensure the message is sent but, this is not the case in Streaming method. In Streaming method, message broker takes the responsibility of sending all the messages without failure. Thus, Delay in streaming method is negligible.

While communication is done using HTTP method, if there is a problem in server application, then requested messages can be lost. The lost messages have to be managed manually and failed messages have to be re-sent. But, in streaming process, if the server is down the messages are stored in queue in the message broker. So, we don't have to worry about the losing the messages. In real world scenario manual management of failed messages are not possible, because it will become problematic in scenarios such as first come first out scenarios.

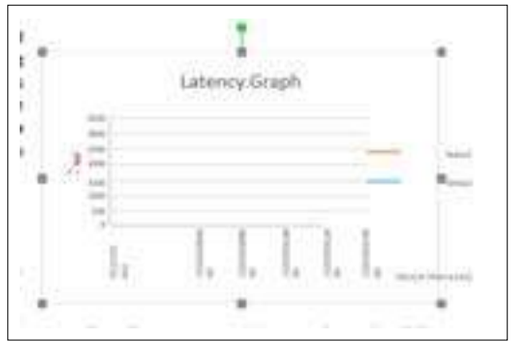


Figure 5. Latency and total time taken for sending 1000 messages

According to Seat reservation system requirement, many people can request for seat reservation at once. In that case request cannot be prioritized manually. Because of the competition to reserve seats, some requests can be lost. We can't handle failed requests without queue. But, streaming method sends all the requests to the queue and messages are stored in the queue until the communication is complete. Thus, no requests will be lost. Once a person requests to reserve a seat that person need to have the minimum credit amount to fulfil the requirement. When the person's credit amount is not sufficient, the request will fail. In this case the chance for reserving the seat should go to the next person who made the request after the first person. This is difficult to achieve by HTTP method. But, data streaming method can fulfil this requirement, because it saves requests in the queue.

Because of the scenario described above, the current developing architecture was carried out using the message

streaming method with queue to avoid any complications. Furthermore, architecture need to enhance for update multiple Query-Models using multiple-queue at same time without update as database level synchronization describe in fig. 1.

Further, according first graph in fig. 6, behaviour of five replicas are the same. But, numerically each curves are different from each other. Second graph in fig. 6 illustrates part of the first graph. It shows that there is some delay between the replicas. That average delay is 3521 milliseconds.

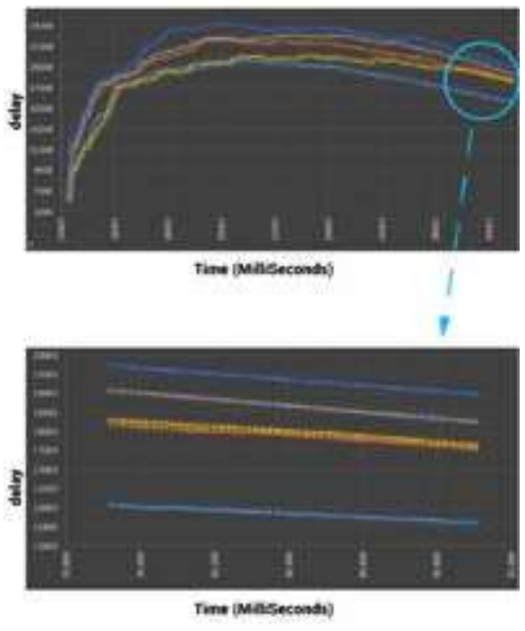


Figure 6 . Behaviour of five replicas

VI. CONCLUSION

This paper contributes to the research on CQRS and event sourcing for Data Handling and Maintaining Data Consistency in Scalable Replicated Microservices. According to the result of this study, we can fulfil below functions. Management of data consistency, replication of unlimited query models and management of eventual consistency. Data inconsistency could happen while communication between command model and query replicas. To prevent that we have to use message streaming process with queues. Predefined queues make limitations for replication of query models and unused queues waste the resources. Generated queues help to reduce wastage of resources. Further, if there is any state change, command model publishes event to all the replicas in same time using streaming, which is better than database level synchronization of replicas.

As the future work, we plan to reduce the average delay to a minimum level using reactive manifesto. Future more, if there is a new instance introduce to the application as a replica of query-model also update as same as other replicas using replaying past states.

ACKNOWLEDGEMENT

For the success of this study it would not have been possible without the kind support, encouragement and help of many individuals and organizations. I would like to extend my sincere thanks to all the parties who joined with me to make this research success. I am highly indebted to Dean Prof. Udaya Rathnayaka, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, and all other staff members of the department of Computing and Information Systems in the Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, for their help and wishes for the successful completion of this research.

REFERENCES

Alshuqayran, N., Ali , N. & Evans, R., 2016. A Systematic Mapping Study in Microservice Architecture. Macau, China, IEEE.

Di Francesco, P., Malavolta, I. & Lago, P., 2017. Research on architecting microservices. Gothenburg, Sweden, IEEE. Fowler, M., 2011. CQRS. [Online] Available at: https://martinfowler.com

Melnik, G. et al., 2012. Exploring CQRS and Event Sourcing A journey into high scalability, availability, and maintainability with Windows Azure. s.l.:Microsoft Developer Guidance.

Pahl, C. & Jamshidi, P., 2016. Microservices: A Systematic Mapping Study. Rome, Italy, s.n., pp. 137-146 . Richardson, C., 2017. Monolithic Architecture. [Online] Available at: http://microservices.io/index.html

The delay could affect the architecture to become vulnerable. Therefore, as a future enhancement it is needed to focus on steps to decrease this kind of delays.

A WEB BASED PAPERLESS MEETING MANAGEMENT SYSTEM

MPL Perera[1] and DU Vidanagama
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]poornimalankani@gmail.com

Abstract- The university system faces many difficulties when scheduling and conducting meetings. The proposed system aims at the design and implementation of a Web-based paperless meeting management system for the university. The system provides a technological solution to utilize such meeting management systems. The system consists of the development of a web-based paperless meeting management system with a data table management option. The proposed system will provide a meeting arrangement may have Kotelawla Defence University (KDU) to get details of the meeting or notify alternatives meeting that can be used through the system. Also, this system provides a meeting room direction map, participants can check meeting details online. Within the day of the meeting attendance mark through Quick Response Text (QR). Provide the special section for out participants to participate in the meeting. To accomplish this, the Web-based paperless meeting management system (WBPMMS) was developed Hypertext Mark-up Language (HTML), JAVA, Android Studio, MySQL, Extensible Markup Language (XML), Cascading Style Sheet(CSS), JavaScript and Web services Description Language (WDSL). The outcome of this study also offers solutions to overcome issues that may occur in the process of manual meeting management. Many industries use QR codes to attract participants and get their information stored in a private database. Further development of the solution will enable of handling the system in android platform.

Keywords- Paperless Meeting management system, Web-Based, Quick Response Text (QR)

I. INTRODUCTION

A typical meeting is led by a chairperson, and minutes which are the recorded deliberations of the meeting are generated at the end. Under the corporate legislation, two classifications of meetings are general meetings and special or extraordinary meetings. These meetings require a quorum, minimum number of members present to make the legally operative. Decisions are taken by the number of votes the assenting and dissenting parties can muster. In the past, before technology played a significant role in our lives, conducting a meeting was a challenging task. The major reason for this weakness was the underdevelopment in the field of telecommunication. Meetings could not be held punctually. Meetings held are not done effectively and efficiently due to poor communication channels among persons. In other words, a meeting with poor ICT interaction was a mere waste of time and energy. Currently, there is no any application software that schedules meetings inside KDU. The traditional method is being used. Up to now, higher management gives instruction to the secretary wants to arrange a meeting. The secretary print meeting minute and distribute or else send an email to staff members regarding the meeting. Therefore, this study focuses in-depth study find solutions for the problems that arise during minute distributions, how the emails are being used to inform the meeting schedules and what if these emails do not either reach the inbox or mistakenly sent to an irrelevant person. Moreover, the problems of displaying the date, time and venue of the meetings on notice boards and most participants do not have an effective impact when they

see the notice displayed. Finally, to find a proper solution to reduce too much paperwork and time consuming that prevail in the present system.

The purpose of the project is to develop a paperless meeting schedule application., we have identified the following objectives.

- Identify the available methods that used for holding a meeting and the problems encountered.
- Identify the available technologies for holding meetings in a paperless manner.
- Design the software solution to address research problems with available technology at present.
- Conduct testing and evaluate and improve the system to meet all functional requirements identified.

II. LITERATURE REVIEW

Through the review, it will be thought about regarding the procedure and available observes in what are the problems of conducting the meeting, technologies as well as what are the existing systems for meeting management.

Meeting has different interpretations and definitions according to the various scholars. Meeting meetings for solving problems or exchanging information it is important to gather for an organization that has analyzed and analyzed all the important decisions (MARUTHAIAH, 2010).

By determination, each meeting the date and place, the system will participate in the system using JavaScript, CSS and MySQL. Using Meeting management system in an organization Can improve an efficient timetable and resources for the meeting (Sultan et al., 2009),(Erik Timmerman and Shik Choi, 2017).

Meeting of most meeting rooms or management system meeting in meeting rooms is basically based on a specific timetable. However, at times during the meetings, there are sometimes times when the meeting halls are frustrating since there is not always a specific date. PIR sensor fusion devices and Ethernet connectivity allows for scheduling meeting rooms and increasing the room utilization of the meeting room (L. D. Tran et al., 2016).

The agent represents each individual multi-representative system representative for one assignment of official delegations to assign delegates to his/her deletants. Multipurpose operators can coordinate their activities and find the solution for their users to meet the needs of their users (Shakshuki and Hoo, 2006),(Yang et al., 2009).

The new hybrid multi-agent architecture tests the troublesome problems that are not generated by the non-programmers to verify the algorithmic functionality of the small representative agents of small representatives of multiple representative agents capable of running on small devices on the mobile device. The algorithm is active and provides the operating system for the hood device (Al-Ratrout et al., 2010).

Meet-me Representatives help mobiles to find time by using cell phones meet me Meeting Planning System The current implementation basically will review the timing of participation with the algorithm with an algorithm. The Meet- me prototype platform for Android developed (Niederer and Schatten, 2009),(Erik Timmerman and Shik Choi, 2017),(K Clark, 2018).

III. METHODOLOGY

A. Approach

In the Analysis of the Meeting Management Problems and web base meeting System process first user should log in to the Web application by providing user's name (email address) and password to the system. The web-based method is suitable for in-house users and Mobile application used for outsiders. All outsiders register with the administrator. The external user will need the link of the android application and the information required by the administrator to his email address. Then the outsider can log in to the meeting through a mobile application. The external person can find out all the information conferring to the meeting (Date, Time, Venue and who will be participating in the meeting). The participants will be selected by the initiator of the event, Meeting minutes also upload by the initiator to the system. According to the number of participants, the system selects the appropriate location for the meeting. And the web application will send a notification for the user whether the information sends successfully or not and give a compliment for the sender for his service Each information that users send

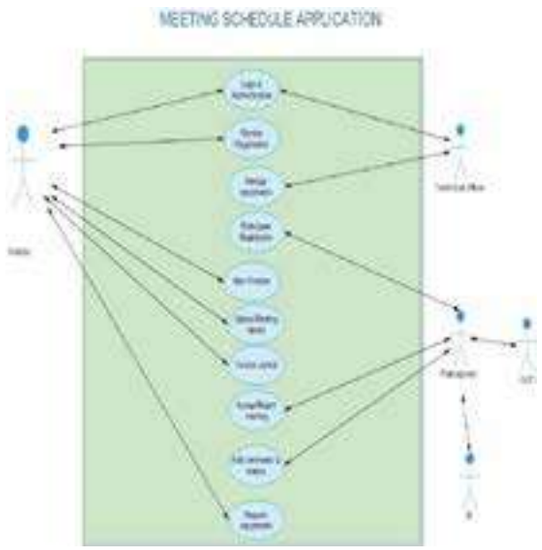


Figure 1. Design module

will stores in MySQL database successfully and for the work with the application it required the continuous internet connection. (Figure1)

B. Technology Adopted

It is necessary to use the new technological methodology for the system. It is very important to use acceptable tools to develop the productive system. Use of any unsuitable tools can solely end up in developing a system with unnecessary errors and faults and use of those badly chosen technologies additionally can end up in crashed when the new system implementation. Badly chosen technologies which can be extremely advanced and complicated will enable manufacturing a system with a top quality, however, these technologies may result in developing a system that spends lots of time and resources to perform a task that is anticipated by the system. It is very important to use an application programming language and the other necessary tools to develop a productive system. Therefore, these technologies and tools can help to develop the system among a minimum development time the most objective of developing this type of an application is to produce the users more efficient work system instead of doing manual approach. Because of that, we should use the most applicable tools available in the market to develop the system. Technological considerations - followed during the development of the system Efficiency and Performance Re-usability and flexibility object-oriented development support so according to the meeting management System java and android studio used to develop the mobile application as well as web application. According to that requirement, the system has developed by using Java and using MySQL database to run on the Windows operating system. This chapter includes the details about the technologies that we are going to use to develop A web-based paperless meeting management system.

- Web Application

  The programming language that is going to apply as the developing language for the system development turned into significantly trusted accuracy, performance. When considering all these technologies which can be associated with the A web-based meeting management system the proposed system can be applied a web-based technology. The spring and hibernate Framework consists of the common language runtime and the Java class library. You can think of the runtime as an agent that manages code at execution time, providing core services such as memory management, thread management, while also enforcing strict type safety and other forms of code accuracy that promote security.

- Database Selection

  Consistent with the above eventualities most of the structures are used the square database to keep facts. It seems it is simple to control and perform. So, the database put in force on the server must able to supply efficiencies operations. Consequently, the proposed system decided on the MySQL server as server. MySQL server is the inspiration of delivering challenge critical performance with in reminiscence technology and quicker insights on any information, whether on-premises or in the cloud, MySQL Server provides an environment used to produce databases that can be accessed from workstations, the Internet, or other media too. Database management or DBMS, store user's data and enables them to transform the information into statistics. Those systems allow users to create, replace and extract facts from their database. A database is an established collection of information. Facts refer to the characteristics of human beings, things and activities. Square server stores every statistic item in its very own fields. In square server, the fields related to a character, thing or occasion are bundled collectively to shape a single

complete unit of records, known as a document. Each record is made up of some of fields. No two fields in a record will have the equal area name.

Throughout an MySQL server database design project, the evaluation of your project wishes identifies all the fields or attributes of interest. If your commercial enterprise desires trade through the years, you outline any extra fields or alternate the classification of present fields.

C. System Architecture.

System architecture is divided into main three layers. They are Application Layer, Presentation Layer and the database Layer. (Figure 2 and 3.).

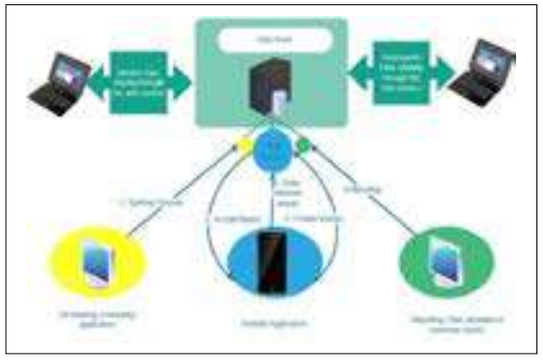


Figure 2. System Architecture

D. Software Design.

There are three user levels in the system. They are Initiator, Participants (external and Internal) and Technical officer. External participants will access the android application. (Figure 3) Admin (Initiator, Internal participants) will access the web application. (Figure 4). For internal participants system will provide the Barcode No (Figure 5).

Figure 3. High-level Rich picture diagram

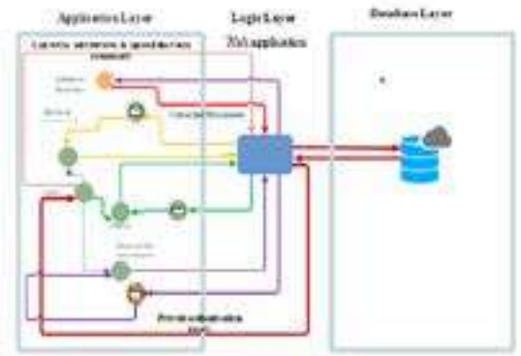


Figure4. Interfaces of Web-Base application (Initiator Window)

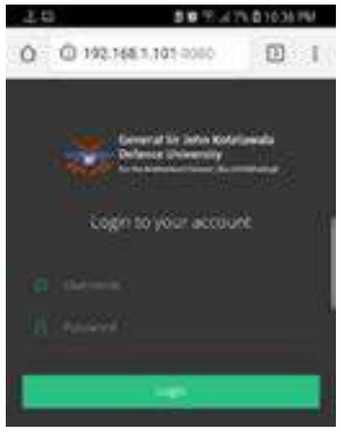


Figure 5. Interface of the web application (Login window)

Figure 6. Interface of the Internal Participants Barcode ID

E. Functional Requirements.

- A "meeting initiator "can update the require documents according to the meeting.
- System should have version control of the documents that update by a "meeting initiator".
- System should have capable to add reviews for the documents during the meeting.
- The system shall keep participants informed about meeting schedules and their changes.
- The initiator can inform users of meeting date & time or else any schedule changes.
- Participants should be able to Inform the attendance to the initiator
- System should be able to provide attendance summary

F. Non-Functional Requirements.

- Efficiency:
  Performance - minimum 512RAM, Space - minimum 1GB
- Security:
  Privacy rules should be enforced, a meeting participant should not be aware of limitations stated by other meeting participants.
  Password protected DB.
- Reliability:
  Meeting locations should be convenient, and information about meetings should be secure.

  A person may not be at two different places at the same time; a meeting room may not be allocated to more than one meeting at the same time.
- Flexibility:
  The meeting date and location should be as convenient as possible, and available as early as possible, to all participants.

IV. EVALUATION

In this, we describe an evaluation of our approach and the developed system while evaluating the objectives achieved how the project deviated from its original specifications and the circumstance identified during the time of the project. This chapter will give the idea of the measure that has been taken to handle the problem occurred and knowledge which have been gathered by supplying solutions for such issues. A determinative evaluation a method for adjudicating the worth of a program while the program activities are forming (in progress). This evaluation is done with the user requirements or the functional requirements. Cumulative evaluation refers to the assessment of participants where the focus is on the outcome of a program. It is done with a high-fidelity prototype to assess the achievement of a product more progressive.

V. CONCLUSION

The results and outcomes engendered in relative to the specificity of the problem domain are enlarged into wider concepts depending on logical assumptions. This chapter aims to clearly accentuate the outcomes and findings of the project and to determine the way of these outcomes and findings can be matched in different contexts that are similar to the problems which are solved by the developed A Web-based paperless meeting management system. In the rest of the chapter, a total summary of the development of the system is given. Furthermore, future enhancements for the developed System have suggested finding out ways to give in additional features to the system and using it outside the business subject in use. The purpose of the project is distributing meeting minutes in the correct manner and get detailed reports. The development team implemented this system in order to determine its ability to satisfy the entire functional and non-functional requirement with special qualities such as flexibility, reliability efficiency and etc., to overcome the drawbacks identified in the system. The study found out that it is feasible to use the language Java and MySQL as database and java in the android studio used to develop the mobile application to develop the project.

It's is a mobile application and web-based project, so this mobile application offers a user to install the application and enter data. This is very helpful for the user to enter the desired information through so much simplicity. The user is mainly more concerned about the validity of the data, whatever he is entering. In Web Server, Initiator provided the option of monitoring the records entered earlier. Data storage and retrieval will become faster and easier to maintain because data is stored in a systematic manner and in a database. Decision-making process would be greatly enhanced because of faster processing

of information since data collection from information available on the computer takes much less time than a manual system.

This system allows getting information about History of the meetings as well as upcoming meeting. This gives efficient and cost-effective. The mobile application can be access by defined user categories.

ACKNOWLEDGMENT

Authors would like to acknowledge the supervisor Mrs.D.U Vidanagama and for her massive guidance, valuable advices, immense assistance and support throughout the research. Special thanks go to all the lecturers of Faculty of Computing whose dedicated support for the project was intensely appreciated.

REFERENCES

Al-Ratrout, S., Siewe, F., Al-Dabbas, O., Al-Fawair, M., 2010.

Erik Timmerman, C., Shik Choi, C., 2017. Meeting Technologies. https://doi.org/10.1002/9781118955567.wbieoc132

K Clark, R., 2018. Meeting Scheduler Using Mobile Agents.

L. D. Tran, A. Stojcevski, T. C. Pham, T. de Souza-Daw, N. T. Nguyen, V. Q. Nguyen, C. M. Nguyen, 2016. A smart meeting room scheduling and management system with utilization control and ad-hoc support based on real-time occupancy detection, in: 2016 IEEE Sixth International Conference on Communications and Electronics (ICCE). Presented at the 2016 IEEE Sixth International Conference on Communications and Electronics (ICCE), pp. 186–191. https://doi.org/10.1109/CCE.2016.7562634

MARUTHAIAH, M., 2010. MEETING MANAGEMENT SYSTEM 24.

Niederer, M., Schatten, A., 2009. Agent-based meeting scheduling support using mobile clients. IEEE, pp. 719–724. https://doi.org/10.1109/ICADIWT.2009.5273904

Shakshuki, E., Hoo, H.-H., 2006. Software Agents for Meeting Scheduler. https://doi.org/10.1109/AINA.2006.315

Sultan, A.B.M., Ghani, A.A.A., Fang, L.S., 2009. Framework for Conference Management System. Comput. Inf. Sci. 2, 90.

Yang, K., Pattan, N., Rivera, A., Griss, M., 2009. Multi-agent Meeting Scheduling Using Mobile Context. https://doi.org/10.1007/978-3-642-12607-9_15

THE RASHOMON EFFECT ON SOFTWARE DEVELOPMENT REQUIREMENT GATHERING PROCESS

I de Silva[1], BMTN Rathnayaka, and M Wickramatunga
Department of Information Technology, Southern Campus,
General Sir John Kotelawala Defence University of Sri Lanka
[1]gifindgif@gmail.com

Abstract- A software project, whatever the size must go through defined stages which known as the Software Development Lifecycle (SDLC). Among five phases of the SDLC the first phase: Requirement Definition needs certain technique to gather requirements and identify functional and non-functional specifications. When conducting a requirement gathering method naturally occurs conflicts. The cause for the conflicts can define as: in single occurrence can be found multiple ideas with different viewpoints. Likely this situation can be called as Rashomon effect. By discussing what are the conflicts occurs when conducting a requirement gathering technique and how Rashomon Effect play on these identified problems, through this paper will discuss about effect and the relationship between Software Development Requirement Gathering Process and Rashomon Effect. Finally, this paper concludes that how to overcome the Rashomon Effect in Software Development Requirement Gathering Process.

Keywords– Rashomon Effect, Requirement gathering process, Interview

I. INTRODUCTION

The Rashomon effect has been defined by Robert Anderson as "the naming of an epistemological framework—or ways of thinking, knowing, and remembering — required for understanding the complex and ambiguous moments" [1] When considering Software requirement gathering, there are different techniques such as; Interviews (one-to-one or group), workshops, Questionnaires, Observation, Feasibility study, Surveys, Document Reviews, Use cases, Prototyping, Current system documentation, Requirements gathering tools like Axia's RFI/RFP Templates, JRD (Joint Requirements Development), Brainstorming. Among mentioned requirement gathering methods here chosen the easiest yet most powerful technique; Interview, because

Interview is a technique which several people will attend to the same event for a certain task. So, the response to the questions asked may be differ from each other. So, the Rashomon effect can be clearly identified.

II. SOFTWARE REQUIREMENT GATHERING PROCESS

This is the main technique used to collect requirements from the end users of a certain system. This is done between the system analyst and the end users of the system. A document which is prepared in technical language be get as the final output. This is one of the most important document which is used by the software development team in any kind of a company. This explains about the main features and the functionalities as functional and non-functional requirements of the system. Simply all the requirements and the expectations of the end user about the system is mentioned through this. Clear communication is the most important thing during this process. Lack of communication may occur misunderstanding between the analysts and the end users. And that will affect to the final output of the system.

Document which is made by gathering the requirements from the user is called as Software Requirement

Specification. Although the document is finally written in a technical language, first the communication is happening with the use of natural language between the analyst and the stakeholder because that will be easy for the understanding of the end user as they may be not very much familiar with the technical languages. The stakeholder should clearly explain their requirements with the analysts as it will affect to the final output of the system. And the requirement explanation on both software and hardware separately is very important.

Software Requirements are mainly two types as Functional and Non- functional. Functional requirements are known as the main functionalities which is in a system and how the system should mainly act through the process. "Functional requirements are the interactions between the system and its environment independent from implementation. Sometime functional requirements are also stated as system constraints (Lauesen, 2002)" [6] These are the main functions which should be accomplished by the system. Non- functional requirements are known as the characteristics of the system which can be used to have a judgement regarding the system. "These requirements define system properties like reliability of the system, response time and storage requirements" [6] This will show the quality of the system to the users.

III. RASHOMON EFFECT

The Rashomon effect occurs when the same event is given contradictory interpretations by different individuals involved in that sense Rashomon effect can be simply defined as a how a group of people react to the same incident in a different way. The reaction of them differ from one to another in the very same incident. This effect was introduced by the movie 'Rashomon' directed by Akira Kurosawa in 1950 and the plot based on a short story named "In a Grove" by Ryunosuke Akutagawa.

Under this observation of any incident is different from each other. This will happen mainly because of different reasons. Absence of proper evidences regarding a certain thing or incident will occur this. Because of lack of the evidences the people related to that incident may have different perspectives regarding that.

The incident, people involved, place, time are in the same event but the story which each participant involved are telling is completely different from each other. There is no connection with story among themselves. Everyone tries to make their own story highlighted among the rest of others. Even if the same story narrated the confessions are differ one to another. This event occurs because of human's multiple ideas with different viewpoints.

IV. WILL RASHOMON EFFECT PLAY ON SOFTWARE REQUIREMENT GATHERING?

As mentioned above in this paper mainly focused on the requirement gathering technique 'Interviews'. Interview is a conversation between more than two people which is happening in an organized way. Most of the time it will happen under a fixed date, time and place. The number of people who are participating to the interview depends on the situation and purpose. The very first thing you have consider before conducting the interview is to identify the nature of the clients which you are going to deal with. The whole other part of the interview depends on that. As a requirement analyst that is a must before the interview.

There are different ways in which the Rashomon effect will play on Interviews when using as a requirement gathering technique.

The pressure given by the society to that incident may also have a lead to this. This is mainly based on the psychological view of the participant. Because everyone cares about how the society look at you in different kind of situations. So, the desires of them may change from one to another with the society they live. Sometimes people avoid telling the truth because of the control and the limits made by the society. That limits may come from their social status, company policies and legal backgrounds or even by the traditional aspects. And if all the participants of that incident want to make themselves to be highlighted they will have different kind of attitude regarding that incident. They will say whatever they want to make themselves important among the rest of participants. This can be identified mainly in lots of companies as many of the employees try to make themselves highlighted among the others and they try to make lots of points to get the attention of the rest. By that the facts they present may be wrong also. Some are trying just to make themselves important and highlighted among others, not to tell the truth.

Interest level on a certain incident may also be a reason for this. The interest level of something can change the content of the whole story. If we have a good interest on the situation the facts, we present to the others are most probably positive. But if we don't have a good intention towards that situation we will not give a good care of what we talk. The facts may be correct or incorrect. Imagine there is an interview with a group of employees, sometimes the information given by them is not what we expected and not much related with the questions asked. So, this may occur because of the lack of interest on that certain topic. And also, some may have a very good intention with this. This all depends on the interest level of the participants.

Lack of communication may also influence on this. One of the main thing that should be clear and confident during an interview is communication. If there is no good communication between the people who participate in the interview problems may arise. If the questions asked were not clear enough the receiver is unable to answer the question in a better way. So that the clients will not give a proper response as they did not get the question in a clear way. Sometimes the answer may not be related to the question and problems may occur. Different participants may answer to the questions asked in a totally different way as they have different communication skills.

Poor communication skills will also have a relation to this. If the participants have no proper communication skills to communicate among them the whole story will be different from each other. The language is very important. All the participants in that communication process should be familiar to a certain language. Both parties should use a language which they are comfortable with. Otherwise the communication may not be successful. And if there is a group of people who can't understand the way they talk with each other the message given through the process is not understandable. So, the communication skill is very much important to get a certain idea without any obstacle.

For an example sometimes after the first interview development team may provide a document which mentioned the requirements given by the client and they will say that mentioned requirements are not the expected and required once. The root of such an incident can be a poor communication.

Mental and physical condition in that moment will also have an effect on this. If the participants of the interview are not in a good mental health at that instance, they will not express the ideas in a better way. Sometimes they will be not in a mood to express their ideas, but they have to do it anyhow because of the rules and regulations of the company. And also, the participants should be physically fit to express the ideas. Otherwise it will badly affect to the whole communication process.

In some interviews the interviewer (narrator) tries to highlight themselves saying the high position of their company, previous and ongoing success projects, how do they complete projects. Those details are not necessary for the interviewee (Client/ End user) because they have come to give a project get done and before coming they already search the basic background information about the company. Giving unnecessary information may lead to boredom of interviewee. In this situation they also do not expose themselves and naturally skip explaining requirements from their viewpoint.

V. POSSIBLE PROBLEMS OCCUR IN AN INTERVIEW AND HOW TO AVOID THEM

Using Observation method, gathered the possible problems occur in an interview and through them figured the possible methods to avoid them. For the observations made, we used some interviews which we had for the requirement gathering process in some of our projects that we have done. To make observations created a constitute interview to gather information to build an Online Leave management system and Online Booking System for KDU Staff Accommodation for KDU Southern. The software development team were act as interviewers and said to gather information needed through interview. Randomly chosen interviewees who directly have connection with the manual leave management system and the staff accommodation booking system were interviewed under different environments and predefined facts (ex: interviewers talking about development team's achievements and capacity to do this work, hold interviews in heavy rainy or humidity environment, use patterns to recognize interviewees who have phobias in patterns) which are directly affects to the software gathering process. As the interviews were done under different people and different situations, the observations done simultaneously when conducting these interviews and figured following problems and by doing discussion with development team got the possible methods to avoid them.

A. Conflicting Requirements

These requirements received from group of end users/ stakeholders or from one stakeholder at different times. Requirements become problematic because they are not match each other. For an example if one stakeholder requires a simple UI and another requires fancy UI. In such cases occurs conflicted requirements and software development team cannot take a straight decision on the problem. This fact clearly effects on identifying Functional and Non- functional requirements of requested system.

Business Analyst can get all stakeholders in the same room and hold a group interview to gather requirements and show them documented requirements in front of them. Then development team can gather requirements without conflictions.

B. Understand the Requirements

Many stakeholders think that they understand the requirements and they said them clearly to the team. But when conducting actual interview, they do not present actual requirements and them self's they do not understand what they need. This lack of understand of the requirements brings problematic moment to the software development team. In other hand if development team did not get the clear idea in requirements result in they will develop a failure system.

The Business Analyst can say to stakeholders list out their requirements before the interview and if there are not clear parts in requirements development team should question them and clear the doubts.

C. Culture and Geographical Area Differences

When conducting interview among different culture and geographical area stakeholders they might be not expose very well to the development team and not say their actual requirements because of the culture, norms or ethics barriers on their own or they ask not matching retirements to be done to the system. Also, some companies may have their own cultures and policies which will limit the exposure to the event. Business Analyst should do a company background checkup before the interview and make sure without violating issues make questioner accordingly.

D. Infinity of Choice

Some stakeholders may present many choices and give hard time to decide what option to select. There might be high percentage of missing important requirements because of the many choices presented and cannot identify the true options needed. These missing facts can affect to the whole system developing process and at the same time less satisfaction to the same stakeholders and decision makers.

A Business Analyst can give limitations to choices and do not give more options to choose.

E. Unclear Feedback

Not having clear understanding of asked questions in interview leads stakeholder to give unclear feedback and they may change their feedback time to time. Result in development team cannot think of options should add to system and decision makers cannot make decisions upon unclear feedback.

This shortcoming can cause because of two parties. One is the development team who make interview questions and those questions may be unclear. The other party is the stakeholder who does not pay attention or not having proper understanding of his own subject area.

As a solution, can suggest that firstly development team need to check out their interview question's clearness and are they directly relate to the particular subject being use in software project. Then they can give an event coverage and inform interviewee to prepare themselves according to the subject. Also, team should select right person to interview. For an example a company needs an automated account system and team should interview people from that department not from HR department even if some of them have known the processes of accounts department.

F. Cliques, Groups and Friendships

Tightly coupled group of individuals or friends may not give requirements properly. Considering a clique or a group they will change their opinions, choices occasionally because of non-stabilized behavior or let arise personal tastes and cause conflicts in information provided. Furthermore, there can be group of stakeholders with sub groups under one company.

Suppose that there is this interviewee group consisting six members; two from HR, two from finance and two from accounts. If HR and finance members are foes each other, they will not communicate each other well even if it is a business-related event. They will surely give different requirements for needed system.

In other hand if they are friends someone with development team they will not take it as a serious task and go out of the topic or think can present whatever they want, change requirements timely and finally development team will filter, wrap up things for them. These above-mentioned facts clearly show giving different ideas for same event.

To minimize this problem development team can observe the situation and say for groups or cliques to pay attention to the business not the personal tastes, ego. In friendship, clear them the idea that works, and friendships are two different things and convince them to give full support to the project in a professional manner.

G. Participation of unlimited number of stakeholders will cause an information overload

When a huge number of clients are participated for the interview, there will be a real mess. So much of people is hard to control. If there are a huge number of participants most of them are not even related to that subject. So that will be a disturbance for the whole process. When there is an unlimited number of stakeholders present in the interview unwanted details and information is presented all the time. And also, some people may not have enough time to express their ideas within the given period of time. That may cause arguments among the participants. Some may express their own ideas which are not even related to the interview. Some may express more than enough. So, the main requirements presented by the stakeholders may hide because of this. The main points which should be highlighted are going down because of the information overflow. Information overflow means the difficulty of understanding the real matters because of the excess of information regarding a certain thing. So, the information overflow may cause bad results in an interview. Because the requirements cannot be clearly identified because of that. When the number of participants in an interview is high both necessary and unnecessary information is collected together. And then an information overload is occurred. So, the requirements cannot be identified clearly among all of them.

As the solution for this problem the number of participants should be limited in an interview. Before the interview is going to hold the clients should be informed about the number of participants that can attend the interview and that will reduce the rush during the interview. And also, the stakeholders should be previously informed about the time period which each of them will get during the interview. That will help not to make any arguments within the interview. When the time duration is previously mentioned the stakeholders can get ready to present the requirements of them within that certain period of time. And also, there may some situations where the interview is included with a huge number of clients with the nature of the interview. So, in such cases, time management is very much important. Everyone should consider more about how much time that they will get during the interview. They should manage the time very carefully. And also, the participants should not be afraid to get the assistance from the people who have effective communication skills.

H. Some people may appear even if the situation is not related to them.

This can be mainly identified in the situations where an interview is holding with the participation of clients in different fields. In this kind of situation some people are trying to express their ideas other than in their own field. As an example, if the situation is gathering of requirements from the HR department, the stakeholders which present the requirements at that moment should be from the HR department. But there are some instances where some people trying to interfere other fields when presenting the requirements. There are instances where the clients in the HR department is presenting their requirements and in the same time some participants from the Accounting department also express their ideas during that. That is a huge disturbance for the clients in the HR department because within that time period they are unable to express their requirements. Even though that time period is not related to them some people try to interfere and express the ideas in other fields. When this kind of situation appears, each field may unable to present their requirements with the given period of time. In this kind of situation some people might say 'this is not my area, but this is how it works'. Even the related clients are available some people try to make their own decisions and that will cause arguments within them. So, this kind of situation will badly affect to find the correct requirements in the relevant field.

As the solution for this problem, the interviews must hold separately with each field at one time. Otherwise some conflicts may arise. If the interview is happening with the participation of the stakeholders from the same field, that can avoid the problems, arguments and the misunderstanding which will arise within them.

I. Feedback from the client may change with environment, time and personal problems

The place where the interview is going to happen is very important because the comfortableness in that environment will lead to a good communication among the participants. But if that environment is not ready to a communication process, a lot of problems may occur. If the environment is too hot or cold that may cause the participants in a bad way. They may not focus in the interview as the environmental condition is bad. And if the environment is too noisy that may also affect to the communication process. If there are problems in the background of where the interview is holding the questions asked and the answers given may not be much clear. If the participants are not satisfied with the environment that they are in, they may not try to give a proper feedback for the questions asked. And the time which the interview is going to happen is also important because different people are comfortable with different time periods. So, if the time is bad the clients may not act to in a better way or may not give a proper feedback. And there may cause another problem. If the clients are running with their own personal problem at that moment, they are not ready for the interview. They may no present their needed requirements clearly because of that. They may physically present in the interview, but they are not mentally ready for the interview. And that will badly affect for the success of the interview. So, the expected results may not come from the participants. So, the whole process and the time will be wasted.

As the solution to this problem the ones who organize the interview should previously consider about the condition of the place where the interview is going to happen. They should choose a place which will be comfortable to anyone. And also, when selecting the clients who are going to participate, they should be informed earlier about the place and the nature of the interview. So, if the participants may not ready with those conditions should be replaced. And also, the participants with bad health and mental condition at that right moment should not attend to the interview. Everyone who are going to attend to the interview should be well informed and ready for the flow of the interview. Sometimes the clients with special needs within that time may appear in the interview. So, the environment must be ready for them too. And also, if any client need any special requirement or have any hardships also should inform to the organizers earlier.

J. Communication Problems

Communication problems are the most critical problem in interviews. This includes: miscommunication, language barriers, wrong assumptions, unclearly defined vocabulary, notation differences. Also, poor communication skills give unclear and ambiguous requirements.

1) Language barriers causes Miscommunication: Unclearly defined vocabulary, subject related words, use of technical words, use of different languages will lead occurrence of misunderstanding and through that stakeholders will do miscommunication.

   Minimizing use of subject, technical words and define vocabulary before interview may help to reduce language barrier in interview.

2) wrong assumptions: Stakeholder's lack of understanding in needed requirements will cause expressing wrong assumptions.

3) Environmental issues: Environmental issues such as noise: conducting an interview in machinery factory or other noisy place or time. And also, other unenviable fact is weather conditions. For an example have to conduct an interview in heavy rainy day, room is not having great sound blocking environment may cause misunderstanding and at the same time miscommunication because cannot hear clearly through the background sounds. Some mental or phycological issues also affect to the communication under the environment. For an example some phobias in patterns: if the environment or room has strips or circles then the interviewee has a phobia in patterns he or she may not give proper response to the questions. Same situation occurs when interviewee do not like some certain colors and the environment have them. These facts will distract the attention of the event.

   As a solution Business Analyst can do an environmental checkup whether the place is suitable for an interview or not, if interviewee have a phobia what is it and find comfortable place to conduct the interview and last but not the least make sure to stay alert on weather forecasting. If weather is not good for interview the alternative solution is not to postpone the interview but to find better place which reduce the weather barriers.

K. Expressing of requirements may change with the managerial level

There are different types of managements in any company who are participating in an interview. The way employees express their requirements change with their management level. When we talk about the high-level management, they are in the superior level of the company. They are the ones who consider more about the fame in their company. Because of that, under some situations they avoid expressing the real requirements to the business analysts. That is because of some reasons. One reason is they have some privacy issues within the company and avoid giving the sensitive data to the outsiders. Because that may cause bad security issues to them. Sometimes the high-level management is not ready to put their company into a lower level by presenting some needed requirements because that shows the company came up to this level without those facilities up to now. That will decrease the demand of the company. So, they avoid revealing the real nature of their needed requirements. That is not a good fact within an interview. And when we talk about the middle and the lower level employees, they do not consider much about the privacy policies of the company as the high-level management. So, they are expressing their requirements very freely and sometimes they exceed their limits. And that will affect badly for the policies in their company. They will give unnecessary information and the requirements may be difficult to identify clearly. So, when the management level changes with their positions, the way they express the requirements about the same thing may be changed.

As the solution for this the interview in a same company should not be done according to different levels of positions. A group should be made with the combination of all levels of employees as the stakeholders. So that will avoid the hiding of some requirements and also the expressing excess information. All the requirements will be presented to a right scale. As every level of positions are included as the clients the requirements can be clearly identified with them. And also, when the stakeholders are selected only within the high-level management some requirements will not appear within them as the low level and middle level employees are the ones who work and move much with the working process of the companies. So, having a combined group with different working levels are very much important when presenting the requirements in an interview.

L. Bad Requirements

Requirements are becoming bad when they are: incomplete, ambiguous, cannot verify. This may lead to an incomplete system. Using a checklist of requirements, their characteristics and test each one of them to ensure all the requirements are good enough to get into work.

VI. CONCLUSION

This paper described the Rashomon effect, Software Requirement Gathering Process and the connection between them. Finally, through findings concluded the consequences of Rashomon effect on interview as a software gathering technique and describe how to overcome the problems occurred.

ACKNOWLEDGMENT

This research was greatly assisted by Academic staff and we thank our colleagues from General Sir John Kotelawala Defence University of Sri Lanka, who provided support to the research.

REFERENCES

Anderson, Robert, "The Rashomon Effect and Communication", Canadian Journal of Communication, Simon Fraser University, Vancouver Canada, 2016. (41(2)): 250–265

Wendy D. Roth and Jal D. Mehta, "The Rashomon Effect: Combining Positivist and Interpretive Approaches in the Analysis of Contested Events", Department of Sociology and Social Policy Harvard University, Sociological Methods and Research, 2002. 31(2):131-173.

Makinster David (2015). A Preface to Rashomon. [Online] Available at: https://www.youtube.com/watch?v=0vVETFktVNQ [Accessed 20 April 2018]

Gordon Briggs (2012). Rashomon in 9 Minutes. [Online] Available at: https://www.youtube.com/watch?v=BP2MhghDal4 [Accessed 22 April 2018]

Shaen McMurtrie (2016). The Rashomon Effect. [Online] Available at: https://www.youtube.com/watch?v=m0ySnhiI1hw [Accessed 22 April 2018]

Umar Sajjad, Muhammad Qaisar Hanif, "Issues and Challenges of Requirement Elicitation in Large Web Projects", Master Thesis Computer Science, School of Computing Blekinge Institute of Technology, Sweden, January 2010

INTERNET OF THINGS BASED FALLS DETECTION AND HEART ATTACK DETECTION SYSTEM FOR ADULTS : SMART WEARABLE

KKVS Anjana[1], RPS Kathriarachchi, and DMR Kulasekara
Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[1]supun94@gmail.com

Abstract- In the new era of communications and technology the Internet of Things (IoT) connects devices, sensors, appliances, people, and things. The IoT can help to enhance the living style of the humans in wider area. Most importantly IoT devices will help to safe guard the living beings ever than before. Most of the elders wish to live independently at homes. Some activity in their daily life is prone to have some accidents, such as falls and heart-attacks. Falls can make people in fatal conditions, even death. Also, Heart attack is a global leading cause of death for both gender of elders and the occurrence is higher than other incidences, the research will target to the adults who are suffering from illness and reduce the death rate of heart attack and reduce the bad effects of fallings early as possible.

The research was conducted within applied research paradigm. The intention was to develop an IoT Based SMART Wearable device. The accelerometer uses to detect the falls and GSM based wearable device is use for the notation purpose. The System will generate automatic call as an alert will be sent to family members with the location. This research also can distinguish condition of people between falls and activity doing daily works. When the system starts monitoring and as soon as patient heart beat goes above or lower a certain limit, the system sends an alert to the elder which then transmits this over the internet and alerts the doctors as well as family member. This helps to determine the problem earlier to reduce the death rate of heart attack.

Keywords- Internet of Things, Wearable device, HIS

I. INTRODUCTION

In the novel span of communication and technology, the explosive evolution of electronic devices, smart phones and tablets can be used for physical or wireless communication has become an indispensable tool for day to day life. The next generation of the connected world is run with the Internet of Things (IoT) that associate devices, sensors, appliances, vehicles, and other "things." Objects or objects may include radio frequency identification (RFID) tags, cell phones, sensors, actuators, and so on. The value of the Internet of Things, make it easy to connect anything, access it anytime, anywhere, and effectively access any service and information about any object. Besides, its outspread the benefits of the Internet through remote control capabilities, data sharing, and continuous connections. Consuming embedded sensors that always open and collect data, and are assured to local and universal networks.

At present, heart disease is the leading cause of death in the elderly. Usually, people with heart disease live at home and ask for medical care when they feel sick. However, they often feel sick at the end of the illness and the disease is ubiquitous. This means that the patient's physical condition should be monitored by the doctor and the doctor will decide when to provide medical care based on the patient's real-time status. A key part of this universal healthcare model is the real-time monitoring system. Using the Internet of Things (IoT) approach, you can monitor important people's functions, no matter where they are and what they are doing. In addition, data collection Ed can be sent to remote physicians at low cost,

which ensures that these experts are constantly aware of the patient's physical condition in real time.

Furthermore, fall of an elderly always lead to serious health issues as the failure of their physical ability. Fracture is the most typical injury in fall of an elderly and there is also a certain possibility to get coma, brain trauma, and paralysis. At most fall situations, the fall process is the main source of injury due to the high impact. However sometimes the late medical salvage could worsen the situation. That means the quicker the salvage comes, the less risk the elderly will face. Progress of technology brings more possibilities to help us protect the elderly. Wearable monitoring devices make it possible with realizing low power components. MEMS (microelectro mechanical systems) sensors have simplified the design and implementation of sensor system. Location based service (LBS) makes it more convenient to trace the elderly in health observing. Beside these, mobile computing makes remote health monitoring easier to apprehend. Several kinds of fall detection methods have been developed or applied in our lifetime. Most of the Systems are for indoor environment, but they are hard to realize in outdoor environment. Motion sensor-based method is also commonly used. Accelerometer could provide motion information directly. The sensor measurements or their proper fusion could be used to extricate a real fall.

In this paper, the author proposed an IoT-based monitoring system and heart attack detection system for pervasive heart diseases healthcare and IoT-based falls detection system for elders.

II. LITERATURE REVIEW

A. Emergency Medical System

(VICENTE et al., 2013) have shown that using ED affords prospect to avoid suffering from care as an alternative to the choices of healthcare substitutes. Furthermore, it is absolutely essential to the ambulance personnel support which influence the elderly patient's caring, consequently enabling her/him to participate in healthcare to the highest possible degree. This is actually accomplished through persistently asking for the elderly person's experiences of health, illness, and suffering.

(Wimalaratne et al., 2017) has said even in Sri Lankan EMS Services are still adolescent and country has to developed in many areas to establish as an effective system. Addition to that, the system of Sri Lankan Healthcare should be developed with providing international post graduate specialist training programs for the relevant authorities. The future of the EMS will depend on the recognition of emergency medicine as a primary sphere, public education campaigns for medical safety, and investment in disaster management system such as medical infrastructure.

B. Existing Systems

(Malan et al., 2004) introduced CodeBlue, a wireless infrastructure that is expected to be used in emergency medical care, integrating low-power, wireless vital sign sensors, PDA and PC-class systems. It will improve the ability of first responders to assess patients on site, confirm the seamless transfer of data between caregivers, and facilitate the effective allocation of hospital resources. In addition, due to the dense network of thousands of devices and extremely unstable network conditions, the infrastructure will support reliable temporary data transmission, flexible naming and discovery schemes, and a decentralized security model. This article introduces our architecture and highlights the research challenges addressed by Code Blue development efforts.

Figure 1: (a) mote-based pulse oximeter. (b) The accompanying patient triage application

(Majumder et al., 2017) have done a periodical development in smart home based remote healthcare technologies. Conferring to the author's smart homes that associate with environmental and wearable medical sensors, actuators, and modern communication and information technology. It facilitates continuous and remote monitoring of the health and well-being of older people on a low cost. This system of smart home countenances old people to stay in their comfortable home. Even though it says a smart home it's not more expensive and has limited medical facilities. But the Medical staff can also track the overall health of older people in real time and provide feedback and support from distant facilities.

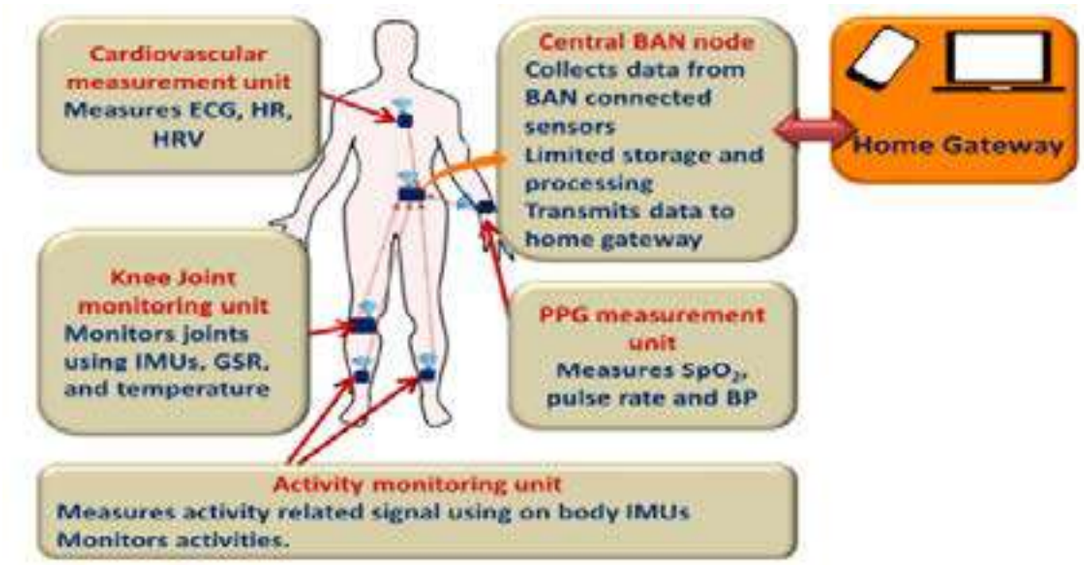


Figure 2: Wireless Body Area Network (WBAN) for wearable medical sensors

In order to achieve all wireless communication sensors and actuators, Standard Protocols for Wireless Sensor Networks (WSN) and Self-Organizing Networks used. Conversely, current protocols considered for WSN do not always apply to WBANs. Figure 2 shows a pictorial representation of a medical WBAN for patient monitoring. Multiple sensors can engaged on clothing or directly on the body, or implanted in tissue, which can facilitate measurement blood pressure, heart rate, blood glucose, EEG, ECG and respiratory rate.

C. Wearable Sensors

(Marschollek et al., 2012) have supposed that ideal sensor systems for health-related parameters. It will be arranged at one point in time and constantly measure and wirelessly report all health-related information. It does not restrict or affect users in any way, nor does it require maintenance. The system similar to the overview of a science fiction novel, but given the effect of technological process. It seems that the system will be reasonable in the not too distant future. This system conflicts with the above requirements and the major problem is energy consumption. Basically it means, that the equipment needs to be charged and repaired frequently. This in turn affects acceptance and compliance. The gap between the harvestable energy and the needs of current sensor systems is still large, but it is shrinking. The demand for service equipment is not only due to energy management but base on the measurement process and the connection between sensors and perceived objects.

(Wolf et al., n.d.), (Koch et al., 2009) have proposed the four axes mobility, connection, measured property and measurement process to organize sensors for health. The related parameters and later improved the scheme (Figure 3).

The presence and availability of wearable sensor systemseasy to deploy and do not burden the patient, if not the main factor impeding the adoption and establishment of health promoting technologies. Currently deployed systems must balance different categories of trade-offs (Marschollek et al., 2012).

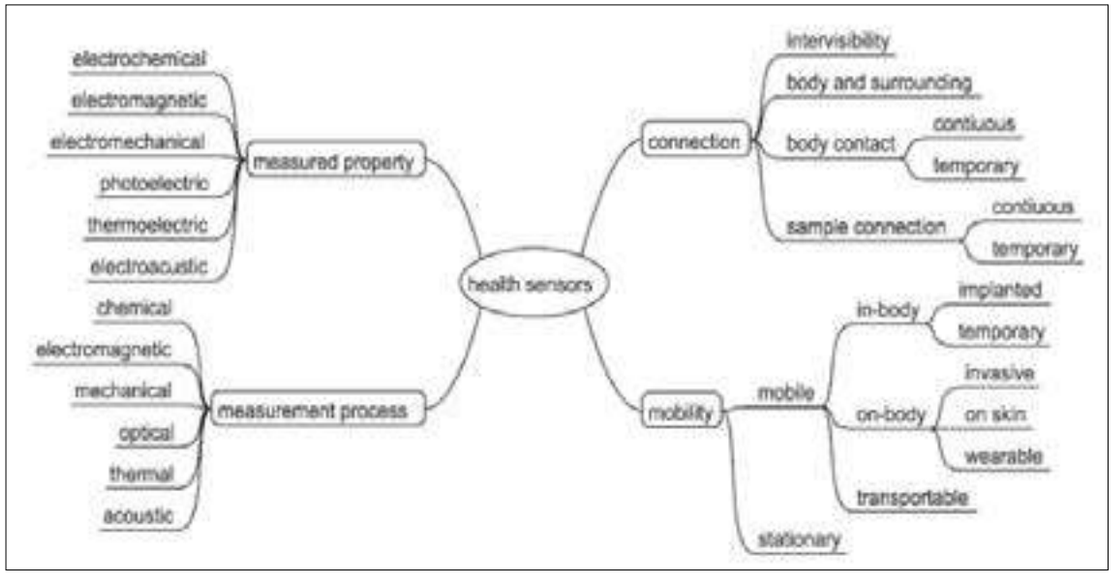


Figure 3: Classification of sensors along four axes: mobility, connection, measured property, and measurement process

(Pan.telopoulos and Bourbakis, 2010) have appraised the current about the development of wearable biosensor systems for health monitoring. Conferring to the biographers, various system implementation methods are associated in the method of identifying the latest technical flaws in current wearable biosensor solutions. this paper displays the architecture of a wearable health monitoring system.(Figure 4).

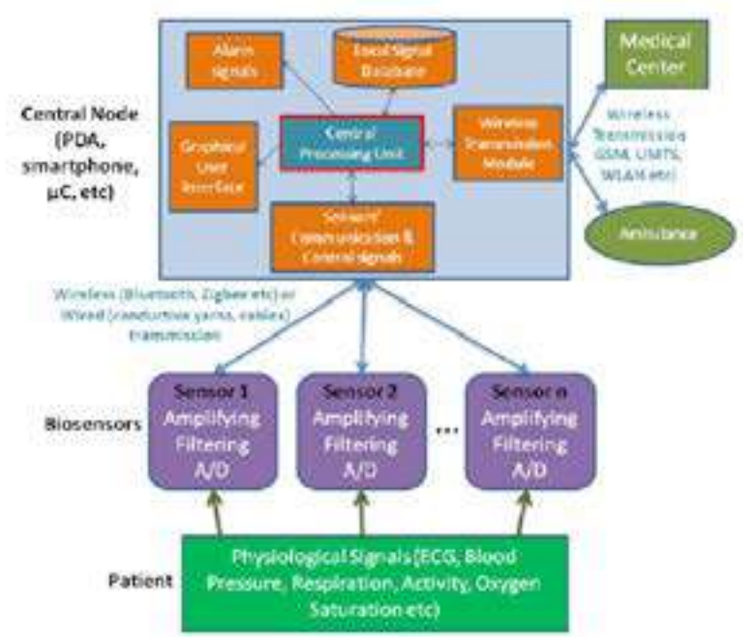


Figure 4 : Architecture of a wearable health-monitoring system

D. Falls Detection Systems

There are several methods for fall detection, such as using a camera, such as Koray Ozcan (Ozcan et al.). Ozcan's research has attached the camera to the body. Therefore, if the direction of the camera changes, it can be concluded that the person has fallen. A very good 86.66% result was obtained from his research. However, some improvements must be considered and many positive errors follow. In addition, accelerometers and gyroscopes are many of the commonly used sensors in today's society. Among them, YanjunLi tried to use an accelerometer sensor (Chen et al.).

His research by using Telos was the chipset that connected to the computer with a wireless connection, but in a small scale. According to that, this detection-fall system is only ideal for indoors.

(Rakhman et al., 2014) has proposed a system utilized a tri-axis accelerometer and gyroscope confined on the smartphones as seen in Figure 5.

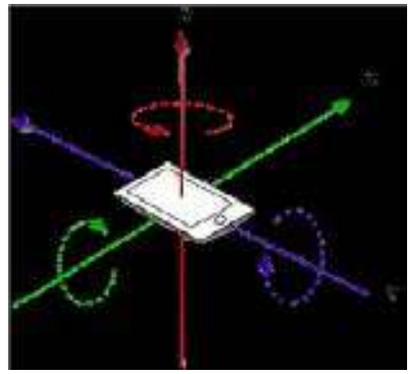


Figure 5: Axis of the gyroscope and accelerometer

(Lan et al., 2009) has presented Smart Fall, which is an automatic fall detection system constructed on subsequence matching, for the Smart Cane system. It uses data from the accelerometers entrenched closely to the handle of the Smart Cane to make implications of current status. (Figure 6)

(Lee et al., n.d.) It is recommended to use the accelerometer of the smartphone to sense the fall detection system of the elderly falling in real time. Mainly its communication spending power in the smartphone to notify the administrator of such an event. The proposed system allows for real-time monitoring of older people who may have fallen. Once the system detects the risk, it uses the smartphone to end the fall and send the location from the smartphone's GPS sensor. Figure 7 shows the real-time location tracking system.

Figure 7 : The real time location tracking system

E. Heart Attack Detection Systems

In a paper presented by Li et al., 2017, an Internet of Things-based heart disease monitoring system is used to popularize health care services. The structure continuously monitors the patient's signs such as blood pressure, electrocardiogram, SpO2 and related environmental indicators. In addition, it offers four different data transfer modes to balance healthcare needs and the need for communication and computing resources. They choose the right equipment to form the sensing layer of the monitoring system. Connectors play an important role in the data transmission of the system. Due to the popularity of smartphones in the system and the openness of the Android platform, the authors focus on using Android smartphones as connectors. The application is responsible for receiving and storing monitoring data from the sensing device via Bluetooth and transmitting the necessary data to give the changed mode of operation. The web-based application was arrested and the doctor could query the monitoring data.

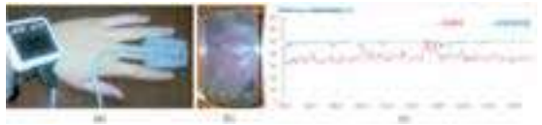


Figure 8 : (a) the SpO2 sensor device; (b) the mobile phone; (c) SpO2 monitoring data.

The figure above shows some of the devices used in the system and examples of monitoring the GUI on the doctor side. Fig. 8(a) is a picture of a sensing device for SpO2 and pulse rate. Figure 3(b) shows the connector in our system, the Android smartphone. Figure 3(c) shows the remote physician's GUI for monitoring patient SpO2 and pulse rate.

Sidheeque et al.) provides a system that uses sensors to detect a person's heart rate using heartbeat detection even when the person is at home. The sensor is connected to the microcontroller, allowing heart rate readings to be checked and transmitted over the Internet. Users can set high and low levels of heartbeat limits. After setting these limits. After the system begins monitoring, once the patient's heartbeat exceeds a certain limit, the system sends an alert to the controller, which then sends the alert over the Internet and warns the doctor and the associated user. In addition, the system will warn that the heartbeat is low. The system also displays the patient's real-time heart rate whenever the user logs in for monitoring. As a result, the person concerned can monitor the heart rate and can immediately get an alarm for a heart attack from the patient and save the patient on time.

In this paper, (Ashrafuzzaman et al.) proposes a system that can detect heart rate using only a camera (Fig. 9) and a commercially available smartphone and using a mobile stethoscope (Fig. 10) to record heart sounds. There are heart attacks and other heart-related diseases. Fuzzy logic is abandoned here, it is part of data mining, and data mining is an expert solution to human diseases. In this technique, the user does not need specific hardware, and he/she can perform measurements almost anywhere in any environment.

III. METHODOLOGY

The research was conducted within applied research paradigm. The intention was to develop an IoT Based SMART Wearable Falls Detection and Heart Attack Detection System which help to reduce the death rates of our adults to and generate automatic call as an alert will be sent to family members with the location if any emergency situation. The research was done focus on a home environment as in the initial stage.

Figure 9:Holding index finger on camera lens

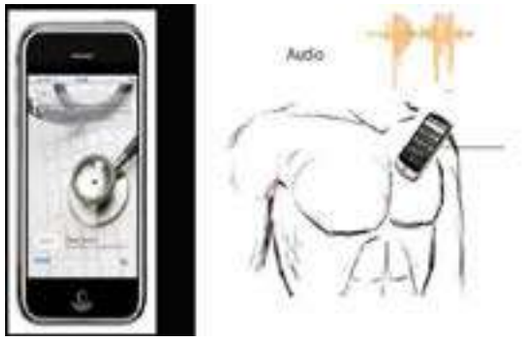


Figure 10: Mobile Stethoscope & Human body with Mobile Stethoscope

A. Falls Detection System

The architecture of the developed system is described in Figure 11. A wearable device is placed on human's hand. The system can detect the elderly's falling by acceleration analysis. Then it will get the elderly's geographic position and send fall alarm short message to guardian(Figure 12). So the elderly who has fallen can get timely help to minimize the negative influence. The authors have added another feature to this device. If the elderly person gets any emergency situation he/she can press the SOS button which authors have added to the device. It sends a call to the guardian.

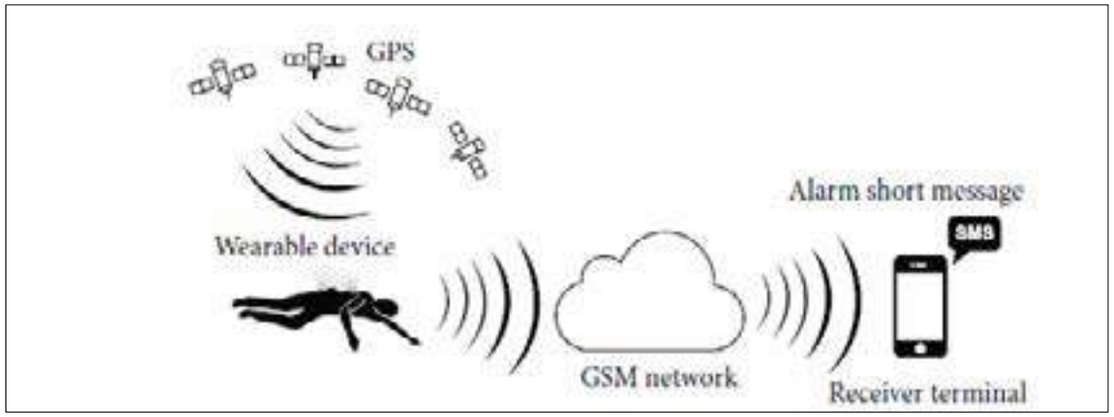


Figure 11: System Architecture

Figure 12: Fall alarm SMS which contains fall location URL

In here authors mainly use sim908 to get the GPS service, GSM function (Figure 13) and accelerometer ADXL345 to get the acceleration. (Figure 14)

Figure 13:SIM 908 module

Figure 14: adxl345

B. Heart Attack Detection System

Figure 15 shows the architecture of the IoT-based monitoring system for heart attack detection system for elders. This is a wearable device and When the elderly person checks his/her pulse rate it shows in a c# application in the computer which the device and the application is connected through WIFI. The person has 2 limitations in the pulse rate which is the lower value and the higher value. When the pulse rate comes near to those values it alerts both the guardian and the doctor through an E-mail. So the doctor can come out for some decisions earlier and the doctor can send medicines which the patient should take via E-mail to the guardian and the patient

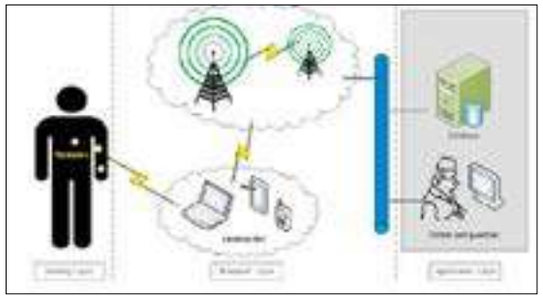


Figure 15 : System Architecture

In here the authors use a pulse rate sensor (Figure 16) for detect the pulse rate and esp8266 WIFI sensor(Figure 17) to connect with the computer.

Figure 16: esp8266

Figure 17: pulse rate sensor

IV. EVALUATION

Summative evaluation was used as the evaluation method to find how the system functions and whether it is up to the expected level to fulfil the users' requirements. At the finalizing stage this evaluation was done to evaluate the product's stability. In summative evaluation, a prototype with most stable build was shown to the user and the feedback was taken to find how far the system is success. The intended target group of the elderly were not employed for testing purposes as a safety measure, since the objective of this project is to design and construct a prototype emergency medical assistance. Therefore, its functionality was tested with the help of colleagues who supported in the testing and evaluation by giving their opinions after using the emergency medical assistance. Several colleagues have given the mobile phones and the computers with the installed C# application with and they were connected to a same Wi-Fi router so the distance from one user to other is low however the when the user initiate the application its worked very accurately and faster than expected. Sometime users send requests same time and its worked fine even that. According to the colleagues used the system they were very satisfied with system. As this project focused on designing and constructing a prototype, comparatively cheap pulse rate sensor and a cheap accelerometer sensor was used. When implementing the actual system, it is recommended to use sensors with high accuracy.

V. CONCLUSION

Increased awareness of the occurrence of falls among the elderly and enrolment of efforts to prevent or diminish such events are highly needed in order to improve the quality of life for elderly people and provide them with convenient fall detection and prevention techniques.

According to the authors, they developed a fall detection system based on a single triaxial accelerometer based wearable device with a SOS button which can use in any emergency situation. The system has low power consumed hardware design which may extend the service time of the wearable device .As normal activity of resting also has similar rotation as falling, it may trigger fall alarm when the body hits ground heavily and alert the relevant people. In future, combining multi-sensing data fusion technology with prediction technologies such as Machine Learning Artificial Intelligence approaches will support developing intelligent fall prevention system based on fall prediction.

A varied of heart attack detection techniques are introduced so far, however they are very expensive and time consuming. Developed a system that measures and detect Human Heartbeat, sends the data to users end by using microcontroller with reasonable cost and great effect. For Human Heartbeat measurement use fingertip, it's in bpm (beats per minute). These calculated rates will have stored in server by transferring through Wi-Fi module via internet. Finally, the stored data in server will be displayed for further analysis by physician or specialist to provide better aid and when the bpm goes to the lower rate and the higher rate it alerts the doctor from the application. From Experimental results, proposed system is user friendly, reliable, economical. Further research work can be carried out for the following issues:

- In Real-time heath monitoring system using ARDUINO can be integrated or implemented in hardware using various types of sensors to detect the human-health conditions of the patients in critical sites continuous Observing of health can be made and the data's will be stored in database.
- In future, a portable Human-Health monitoring system can be designed using Arduino.

ACKNOWLEDGMENT

It would not have been conceivable without the generous support and help of numerous Individuals who were there throughout the project. The Author might want to expand thanks for every one of them. Author would like to express sincere gratitude to the supervisors of this project Mr.RPS Kathriarachchi & Mr. D Maithri Ranga Kulasekara for all inspiration and direction to do this research and kind support and assistance given throughout the research paper writing.

REFERENCES

Icente, V., Castren, M., Sjöstrand, F. And Wireklint Sundström, B. (2013). Elderly patients' participation in emergency medical services when offered an alternative care pathway

Wimalaratne, K., Lee, J.I., Lee, K.H., Lee, H.Y., Lee, J.H., Kang, I.H., 2017. Emergency medical service systems in Sri Lanka: problems of the past, challenges of the future. Int. J. Emerg. Med. 10. https://doi.org/10.1186/s12245-017-0127-x

Malan, D., Fulford-Jones, T., Welsh, M., Moulton, S., 2004. Codeblue: An ad hoc sensor network infrastructure for emergency medical care, in: International Workshop on Wearable and Implantable Body Sensor Networks. Boston, MA;

Majumder, S., Aghayi, E., Noferesti, M., Memarzadeh-Tehran, H., Mondal, T., Pang, Z., Deen, M., 2017. Smart Homes for Elderly Healthcare—Recent Advances and Research Challenges. Sensors 17, 2496. https://doi.org/10.3390/s17112496

Marschollek, M., Gietzelt, M., Schulze, M., Kohlmann, M., Song, B., Wolf, K.-H., 2012. Wearable Sensors in Healthcare and Sensor-Enhanced Health Information Systems: All Our Tomorrows? Healthc. Inform. Res. 18, 97.

Wolf, K.-H., Marschollek, M., Bott, O.J., Howe, J., Haux, R., n.d. Sensors for health-related parameters and data fusion approaches 9. Pantelopoulos, A., Bourbakis, N.G., 2010. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 40, 1–12. https://doi.org/10.1109/TSMCC.2009.2032660

K. Ozcan, A. K. Mahabalagiri, M. Casares, and S. VeJipasalar,"Automatic Fall Detection and Activity Classification by a Wearable Embedded Smart Camera," IEEE J Emerg. Sel. Top. Circuits Syst., vol. 3, no. 2, pp. 125-136, Jun. 2013.

Y. Li, G. Chen, Y. Shen, Y. Zhu, and Z. Cheng, "Accelerometer-based fall detection sensor system for the elderly," 2012, pp. 1216-1220

Rakhman, A.Z., Nugroho, L.E., Widyawan, Kurnianingsih, 2014. Fall detection system using accelerometer and gyroscope based on smartphone. IEEE, pp. 99–104. https://doi.org/10.1109/ICITACEE.2014.7065722

Sidheeque,A., Kumar, A., Balamurugan,R.., Deepak, K .C.,Sathish,K.,2017. Heartbeat Sensing and Heart Attack Detection using Internet of Things: IoT

Ashrafuzzaman, Huq, M., Chakraborty, C., Khan,R.M., Tabassum,T., Hasan,R. 2015. Heart Attack Detection Using Smart Phone. INTERNATIONAL JOURNAL OF TECHNOLOGY ENHANCEMENTS AND EMERGING ENGINEERING RESEARCH, VOL 1, ISSUE 3 23 ISSN 2347-4289

LIST OF REVIEWERS

Snr Prof AS Karunananda, UOM

Prof K Pulasingha, KDU

Prof Prasad Jayaweera, USJP

Dr Hesiri Dhammika Weerasinghe, UOK

Dr HL Premarathna, UCSC, UOC

Dr Ruwan Wickramacrachchi, UOK

Dr SGVS Jayalal, UOK

Dr (Mrs) DDM Ranasinghe, OUSL

Dr (Mrs) KGHUW Ratnayake, OUSL

Dr SR Liyanage, UOK

Dr (Mrs) ATP Silva, UOM

Dr (Mrs) Chethika Abeynayaka, UOM

Dr Sagara Sumatipala,UOM

Dr Asanka Dharmawansa, WUSL

Dr (Mrs) Manjula Maduwanthi, ESoft

Dr Chamara Liyanage, Japan

Dr Salinda Premadasa, UOS

Dr L Ranathunga, UOM

Dr Sachintha Pitigala,UOK

Dr (Mrs) Wijendra Gunathilake, KDU

Dr LP Kalansooriya, KDU

Ms Piyumi Ishanka, SUSL

Mrs M Wagaarichchi

Mr Kasun Jinasena, USJP

Mrs AMCH Attanayake,UOK

Mrs DU Vidanagama, KDU

Maj RMM Pradeep, KDU

Mrs N Wedasingha, KDU

Mr PPNV Kumara, KDU

Mrs SCM Sirisuriya, KDU

Mr RPS Kathriarachchi, KDU

Mrs WPJ Pemarathna, KDU

Mr B Hettige, KDU

Mr ADAI Gunasekara, KDU

Col SPP Pakshaweera

Mrs WGDM Samankula, UOK