The Potential of One-Shot Learning for Drug Discovery – A Review

K Anuradha[1#], DPM Lakshan[1] and WAAM Wanniarachchi[1]

[1]*General Sir John Kotelwala Defence University, Sri Lanka*

[#] kanuradha@kdu.ac.lk

Abstract: *Compounds with specific chemical properties for treating diseases are sought through drug discovery. The search for drugs can be made more efficient, less expensive, and less time-consuming by incorporating automation. New approaches and technologies in drug discovery have grown dramatically over the past few decades. "One-shot" learning is the best hope for the widespread adoption of machine learning in all industries. In this work, we show how one-shot learning can reduce the amount of data required to make meaningful predictions in drug discovery applications. With Few-Shot Learning (also referred to as One-Shot Learning), models can be trained to learn the desired goal with less data, like humans. The study's objectives are to explore the most prominent ways to identify and forecast drug discovery, potential applications, and several remaining challenges. Chemical structures can be represented using some structural descriptors, a similarity measure is used to compare them, and a strategy can be used to predict the activity of a query compound in this manner. This review will serve as an impetus for future experiments that seek to validate the use of one-shot learning in the chemical sciences.*

Keywords: *One-Shot Learning, Few-Shot Learning, Drug Discovery, Machine Learning*

1. Introduction

For the purpose of developing drug candidates, various properties of compounds, including efficacy, pharmacokinetics, and safety, must be simultaneously optimized in order to make small-molecule drug discovery a difficult multidimensional problem. In the traditional way of drug development, scientists test millions of molecules but only a few make it to preclinical or clinical testing. It is long, expensive, and has a high failure rate. There are many ways to reduce the complexity of drug discovery and avoid high costs and time spent bringing medicine to market, such as embracing new automated technologies. For many years, the drug discovery industry has relied on automation. Pharmaceutical companies will be able to make better decisions more quickly thanks to automation in drug discovery. It's no longer enough to use automation only for high-throughput screening. Automated systems in drug discovery have been around for a long time, despite the perception that they are new. They were first used in pharmaceutical companies to conduct high-throughput screening (HTS) experiments. First, automation was implemented in laboratories to help with high-throughput drug development and production. More recent technological advances have made it possible to improve precision, compliance, and reproducibility even further.

The Search for New Drugs Medical research is all about scientists. New developments in automation may help streamline some lab processes and allow scientists to concentrate on their core competencies. In the laboratory, automation increases efficiency, reducing waste, and speeding up the drug discovery process. The world's most brilliant medical minds are working nonstop to discover new treatments and cures for the world's most deadly and perplexing diseases. The ability of automation to perform large volumes of repetitive tasks has revolutionized the laboratory environment, relieving scientists of some of the modern-day pressures and increasing reliability, throughput, and reproducibility.

As a result of pioneering laboratory work over the past few decades, polio has been nearly eradicated and breakthrough drugs have been developed to combat HIV and cancer. The field of machine learning has made several significant advances in recent years that address complex issues. Machine learning and deep learning frameworks require a lot of data in order to have a practical impact on drug discovery. The problem is that there isn't always enough information for every task. Because of this, the need for machine Learning Algorithms that can learn from a smaller number of examples has become more pressing as Machine Learning enters mainstream domains like healthcare. One-shot learning refers to the idea of learning a class of objects from a small amount of data.

It is a challenge to learn new classes with a small training dataset of just one or a few images per category, known as one-shot learning (OSL). In order to distinguish one-shot learning from standard single object and category recognition algorithms, it relies on the transfer of prior knowledge from previously built models. The ability to learn with fewer training examples is made possible by the transfer of knowledge. When a model has been trained using large data sets, it can be used to perform the same task with less training data. This is known as knowledge transfer.

When one-shot learning is used in conjunction with deep learning, metric learning is almost always involved. FSL should generalize if the transferred information is accurate. By predicting the category of each pair of examples separately instead of performing a direct prediction for each object, a separate training set with sufficient examples for each category can be generated. Meta-learning, on the other hand, is used in a variety of ways to overcome the challenge of learning from a limited number of examples. A new

approach is presented in this paper that discusses and demonstrates how one-shot learning can significantly reduce the amount of data needed to make meaningful predictions in drug discovery applications.

2. Literature Review

The paper by Mahdi Rezaei and Mahsa Shahidi describes Zero-shot learning and its applications in autonomous vehicles and COVID-19 diagnosis(Rezaei and Shahidi, 2020). Besides the barrier of a limited annotated data set, they have performed a comprehensive and multi-faceted review of the Zero-Shot/Generalised Zero-shot Learning challenge, its fundamentals, and variants for various scenarios and applications such as COVID-19 diagnosis and autonomous vehicles. There are four space-wise embedding categories for the most recent state-of-the-art methods. In order to solve ZSL, they've analyzed the most popular datasets and their corresponding subsets. It also contributed to the results of some common baselines and elaborated on assessing each group's advantages and disadvantages as well as the ideas behind different areas of solutions to improve each group. Their evaluation shows that data synthesis methods and combinational approaches have the best results, as by synthesizing data, the problem shifts to the classic recognition/diagnosis problem, and by combining other methods, the model utilizes the advantages of each embedding technique.

In the paper "Machine Learning in Drug Discovery: A Review," ML models can perform alternative tools like PPT inhibitors and macrocycles with traditional algorithms if they are trained in a known framework i.e., the compound structure. Because of their high success rate in clinical trials, chemical structures and QSAR models derived from pharmaceutical data can also be included in deep learning models(Dara *et al.*, 2022). Artificial Intelligence has taken a significant step forward in computer-aided drug development, according to the authors. The use of machine learning techniques in drug discovery has run into some problems. Because multiple deep neural networks are trained effectively on a large volume of data, the performance of deep learning methods can directly influence data mining innovation. The primary goal is to address the issue of automatic transfer learning. Another problem with neural networks is that many parameters can be tweaked during training, but some theoretical and practical frameworks are inaccessible.

Introducing the task of low data learning for drug discovery, "Low Data Drug Discovery with One-Shot Learning" provides an architecture for learning such models. Using this architecture, they have shown that it outperforms other methods for learning with low data sets. They found that on the Tox21 and SIDER datasets, one-shot learning methods outperform simpler machine learning baselines in terms of performance(Altae-Tran *et al.*, 2017). These findings are of particular interest to the SIDER collection, which is comprised of high-level phenotypic side effects observations. According to them, their work is not simply an adaptation of one-shot learning to molecular data sets but rather an entirely new approach. Previous one-shot algorithms18 attempted to perform object recognition for new classes of images using only a small number of examples from each class as a learning tool.

"One-shot Learning Approach for Unknown Malware Classification" takes advantage of these advancements to introduce a novel method to help malware analysis identify and categorize malicious software in less than 10 known samples(Tran, Sato, and Kubo, 2018). Using natural language processing and a Memory Augmented Neural Network (MANN) for a one-shot learning task has been demonstrated in this paper. Even with only one recognized sample, classification accuracy is quite good.

"Wheat Disease Recognition through One-shot Learning Using Fields Images" proposes a wheat disease recognition network based on one-shot learning that not only requires a small number of images for training but can also accommodate new categories because it can be trained even on a few images of a new type(Mukhtar *et al.*, 2021). By providing a few images of diseased plants, farmers can quickly re-train their network and begin testing again. As a feature extractor, they've used the MobileNetv3 network, which is extremely fast and accurate.

In Nikhil Thakurdesai's paper, one-shot learning is used in another way. Different face recognition algorithms have been proposed, but one stands out in the case of a small dataset: one-shot learning.' To learn something in "one shot," all you need is one piece of training. In order to address this issue, a solution is presented in this paper(Thakurdesai, Raut, and Tripathi, 2018). Large datasets are required by neural networks to achieve high accuracy. Using only one training sample, this paper proposes a solution to the problem of low face recognition accuracy by reducing the number of training samples needed to one.

3. One-Shot Learning Approach

It is a classification task where one or a few examples are used to classify many new examples in the future. One-shot learning is Classification tasks that are carried out using past data by one-shot learning. Many machine learning applications can be found in situations where data is scarce, and this technology can help. There's a striking resemblance between the way humans learn and recognize new concepts without having previously seen them. In one-shot learning, we only have a single example of each class. Now the task is to classify any test image to a class using that constraint. Multiple training examples are used in the form of few-shot learning, a more flexible version of one-shot learning. An approach to model building that relies on the transfer of knowledge from other models to use a training set of only a few examples was coined in 2006 by Fei-Fei Li et al. in the field of computer vision as the term "one-shot learning." When a model has been trained using large data sets, it can be used to perform the same task with less training data. This is known as knowledge transfer. The

challenge of learning new classes from a small training dataset of one or a few images per category is known as "few-shot learning." FSL should generalize if the transferred information is accurate. Meta-learning, on the other hand, is used in a variety of ways to overcome the challenge of learning from a limited number of examples. Overfitting is a common problem, so the main challenge is to improve generalization.

4. Methodology-One-Shot Learning For Drug Discovery

The Search for New Drugs Medical research is all about scientists. To develop drug candidates, various properties of compounds, including efficacy, pharmacokinetics, and safety, must be simultaneously optimized to make small-molecule drug discovery a difficult multidimensional problem. The world's most brilliant medical minds are working nonstop to discover new treatments and cures for the world's most deadly and perplexing diseases. Large-scale testing is now possible thanks to automation. It has also lowered the cost of electricity and other equipment. Compounds with specific chemical properties for the treatment of diseases are sought out through drug discovery. The approach used in this search has become increasingly important in computer science recently, as machine learning techniques have exploded in popularity since they became more accessible. To speed up the research process and lower the costs and risks associated with clinical trials, this review identifies relevant literature on one-shot learning techniques for the discovery of new drugs.

Figure 1. Various fields in drug discovery by using Machine Learning
Source: Author

Large amounts of manually annotated data ("labeled data") are frequently required by machine learning models. That means that they need constant human intervention to retrain the underlying model with more data. One of the most pressing issues in machine learning is how a model can learn if only a few samples are available instead of using too much data. Deep learning and machine learning are typically more accurate when more data is used. Computational drug discovery relies heavily on the use of similarity searches. Chemical structures can be represented using some favourite descriptors, a similarity measure is used to compare them, and a strategy can be used to predict the activity of a query compound in this manner.

As in the case of one-shot learning, this procedure can be used even if only one or a few examples of each of the activity categories are known. Both methods have their advantages and disadvantages, but there is a major difference between the two. The primary difference is that the descriptors, the similarity metric, and the comparison method all have to be learned from scratch in the case of one-shot learning. Medicinal chemists have grown confident in the similarity search over the years because of its universality, efficiency, conceptual simplicity, and high level of trust. Cons: It's a heuristic approach that's rigid, which leads to suboptimal results. But one-shot learning's greatest advantage is its ability to learn from examples while also finding the best solutions for specific datasets. A few drawbacks of this approach should be considered as well. These include using black-box approaches that are difficult to interpret and therefore not well understood by most pharmaceutical scientists. Another drawback is that the criteria used to select auxiliary data for knowledge transfer are unclear because this approach relies on poorly understood black-box approaches that aren't well understood by pharmaceutical scientists. As a result, we can see that one-shot learning has a good chance of becoming a viable option for drug discovery with further exploration and development.

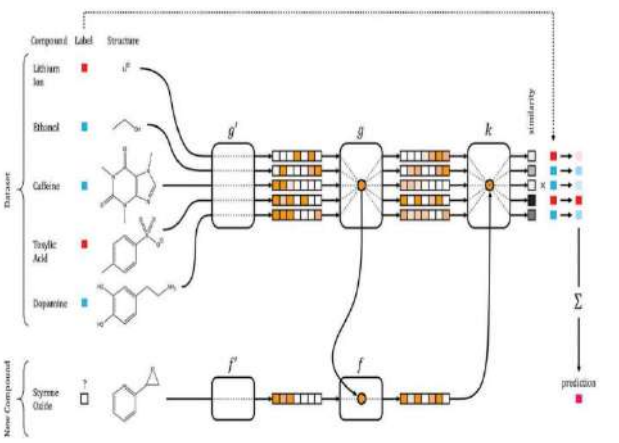

Figure 2. The IterRefLSTM (Iterative Refinement Long Short- Term Memory) architecture for drug discovery one-shot learning. The neural network's augmented memory is preloaded with feature vectors representing a labeled reference set of chemical compounds (left). An attention mechanism weighs the most similar compounds more heavily than the test compound's feature vector (bottom) (right). The activity class of the test compound can be predicted using the weighted combination of all labels.

Source: https://www.tandfonline.com/doi/full/10.1080/17460441.2019.1593368#:~:text=Several%20benchmark%20studies%20reported%20in,adverse%20effects%20of%20marked%20drugs.

i. LIMITATIONS OF ONE-SHOT LEARNING

One-shot learning has some drawbacks, despite its allure. Only one Siamese neural network is capable of solving a given problem. Facial recognition neural networks can't be used for tasks like determining if two photos are of the same dog or cat because they're designed for one-shot learning. Other variations affect the neural networks' sensitivity. For example, if the person in one of the images is wearing a hat, scarf, or glasses, but the person in the other image isn't, the accuracy can suffer significantly. The process of discovering new drugs is extremely complex. There is a risk in using one-shot learning techniques in drug discovery because accuracy is so critical in the medical arena.

5. Discussion

Low data learning for drug discovery is described in this paper, along with an architecture for creating such models. The pharmaceutical industry uses AI technology, including machine learning algorithms and deep learning techniques, on a daily basis. Images and omics data have been the source of many problems for ML techniques in the pharmaceutical and health care industries. The fact that one-shot learning can do well in these predictions is a strong indication that these methods could offer strong performance on small biological data sets, given the amount of uncertainty in these predictions. Only through experimentation can one-shot learning in chemistry be proven to be effective.

References

Altae-Tran, H. *et al.* (2017) 'Low Data Drug Discovery with One-Shot Learning', *ACS Central Science*, 3(4), pp. 283–293. Available at: https://doi.org/10.1021/acscentsci.6b00367.

Dara, S. *et al.* (2022) 'Machine Learning in Drug Discovery: A Review', *Artificial Intelligence Review*, 55(3), pp. 1947–1999. Available at: https://doi.org/10.1007/s10462-021-10058-4.

Mukhtar, H. *et al.* (2021) 'Wheat Disease Recognition through One-shot Learning using Fields Images', in *2021 International Conference on Artificial Intelligence (ICAI). 2021 International Conference on Artificial Intelligence (ICAI)*, pp. 229–233. Available at: https://doi.org/10.1109/ICAI52203.2021.9445266.

Rezaei, M. and Shahidi, M. (2020) 'Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review', *Intelligence-Based Medicine*, 3–4, p. 100005. Available at: https://doi.org/10.1016/j.ibmed.2020.100005.

Thakurdesai, N., Raut, N. and Tripathi, A. (2018) 'Face Recognition using One-shot Learning', *International Journal of Computer Applications*, 182(23), pp. 35–39. Available at: https://doi.org/10.5120/ijca2018918032.

Tran, T.K., Sato, H. and Kubo, M. (2018) 'One-shot Learning Approach for Unknown Malware Classification', in *2018 5th Asian Conference on Defense Technology (ACDT). 2018 5th Asian Conference on Defense Technology (ACDT)*, Hanoi: IEEE, pp. 8–13. Available at: https://doi.org/10.1109/ACDT.2018.8593203.

Author Biography

Anuradha Kurunayaka is a Temporary Instructor at the Department of Information Technology, Faculty of Computing, General Sir John Kotelawala Defence University. She obtained her BSc (Hons) in Information Technology from the same university. She is a former Associate Operations Analyst. Her current research interest is Machine Learning, Data Mining, and Information Systems Management.

DPM Lakshan is a Temporary Instructor at the Department of Information Technology, Faculty of Computing, General Sir John Kotelawala Defence University. He obtained his BSc (Hons) in Information Technology from the same university. He is a former Associate IP Network Engineer. His current research interests include Machine Learning, Big Data, and Networking.

WAAM Wanniarachchi is a Lecturer (Probationary) at the Department of Information Technology, Faculty of Computing, General Sir John Kotelawala Defence University. He obtained his BSc (Hons) in Information Technology from the same university. He has Certificate of Teaching in Higher Education, General Sir John Kotelawala Defence University, and a former Associate Business Analyst. His current research interests include Machine Learning and Optimizing Higher Education.