

# Bacteria Identification and Classification in Drinking Water Using Image Processing Techniques and Machine Learning

GUDU De Silva, WMKS Ilmini

*Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka*

**Abstract.** Drinking water considers as a major requirement for all the animals in this world. But what happens when it becomes polluted? While we can consume purified water by traditional purification Systems why we cannot assure that there can be no threat or harm to public health? The people in rural areas in Sri Lanka are still adapted for traditional reservoirs like rivers, waterfalls, ponds, water streams that may vulnerable to waterborne pathogens which are sometimes harmful (pathogenic bacteria). In this paper, we analyze more about the bacteria that can identify in drinking water like E. coli (also known as Escherichia coli), Salmonella Enterica, Shigella, Legionella Pneumophila, etc. Some bacteria can cause common severe infections like diarrhea, vomiting, loss of coordination, muscle ache, fever but it also can be susceptible to people with bad health conditions which can end up with death. Microbiologists who conduct tests at labs should well expertise with experience to make them accurate and efficient. During the classification of bacteria, it takes much more time and the results may not quite right. Conventional software Systems use mathematics for image analysis algorithms which is hard to adapt, so here we focus on Machine Learning with Image processing techniques for clear identification and classification of bacteria in drinking water. Bacteria Identification and classification experiments are yet trivial and still have a concern on new methods to increase productivity. According to some undergone research works color of microorganisms is the key feature to improve the accuracy of recognition when the shape may appear similar. The most common algorithms used for the classification of bacteria can be dictated as support vector machine, K-nearest Neighbors and Random Forest algorithms.

**Keywords:** *Classification, Feature Extraction, Bacterial Morphotypes, Convolutional Neural Networks, Supervised Learning, Unsupervised Learning*