

**A MACHINE LEARNING APPROACH FOR  
DEPRESSION DETECTION IN SINHALA-ENGLISH  
CODE-MIXED LANGUAGE**

**Muhammed Fahd Fazal**

A dissertation submitted in partial fulfilment of the requirement for Bachelor of  
Science (Honours) degree in Computer Science

**School of Computing  
Informatics Institute of Technology, Sri Lanka  
in collaboration with  
University of Westminster, UK**

**2023**

## Abstract

The most significant and prevalent mental illness in the world today is depression. In Sri Lanka, most depressed people won't recognize the illness in its earliest stages even though it is a treatable condition. Therefore, the rate of suicidal behaviour has been rising quickly. Since social media is popular and easily reachable to anyone, most of the Sri Lankan people tend to share depressive feelings and thoughts intentionally or unintentionally even personal matters on social media hoping to get attention from others without properly consulting a mental healthcare professional to treat their mental health.

As there is a difficulty in early depression detection among Sri Lankan people, the proposed system will be a solution to it by detecting depression based on textual contents on social media which are written in Sinhala-English code-mixed language. This is a research project based on natural language processing and machine learning. ExtraTreesClassifier which is an ensemble machine learning algorithm was used in this system to classify Sinhala-English code-mixed text into depressive or non-depressive class accurately.

End users of this proposed system would be mental health professionals who will detect depression and help people. The proposed system was tested with live data collected from social platforms and it achieved an accuracy of 79.13% in detecting depression.

**Keywords:** Depression Detection, Machine Learning, Natural language Processing, Classifier, Sinhala-English code-mixed language, Singlish

### Subject Descriptors:

- Computing methodologies → Machine learning → Machine learning approaches → Classification and regression trees
- Computing methodologies → Machine learning → Learning paradigms → Supervised learning → Supervised learning by classification
- Computing methodologies → Artificial intelligence → Natural language processing → Information extraction
- Applied computing → Law, social and behavioral sciences → Psychology