



INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Sinhala Handwritten character recognition and segmentation

Key Words – Deep Learning

A Project Proposal by

Mr. Tharinda Widanage

W1742258- 2018588

Supervised by

Mrs. Dileeka Alwis

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering degree at the University of Westminster.

April 2023

Abstract

In order to facilitate interactions between people and computers, technology has been enhanced to recognize both printed and handwritten materials. But there is still a lot of debate over Asian languages. Since Sri Lanka is the one country in Asia that speaks Sinhala just like its official language, character recognition for that language is currently a work in progress.

Sinhala characters differ from those in other widely spoken languages by having linkages between characters in a circular, complicated pattern. The Sinhala language uses 60 fundamental characters (non-cursive). A Sinhala character set is further improved by modifiers that can be applied to basic characters. As a result, it can be challenging to identify Sinhala handwritten characters.

Most character detection and segmentation research efforts employ patterns match and image processing approaches. On the other hands, these approaches are unable to adapt to the changes.

This study aims to develop a handwritten Sinhala character identification system using convolutional neural networks (CNN). The main objective of this study is to create a system that can accurately and reliably recognize Sinhala handwriting using a combination of classification and segmentation techniques. This study generated a dataset of 200 samples each of the 31 character classes. This dataset was used to train a CNN model, which has a 98 percent overall accuracy rate.

Key Words: Deep Learning, Sinhala Handwritten Character Recognition, Convolutional Neural Network, and Image Processing