

# **SINHALA HANDWRITTEN CHARACTER RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK**

**Magina Arachchige Samasha Nayananthi**

A dissertation submitted in partial fulfilment for the requirement for Bachelor of  
Engineering (Honours) degree in Software Engineering

**Department of Computing**

**Informatics Institute of Technology, Sri Lanka  
in Collaboration with  
University of Westminster, UK**

**2021**

## Abstract

To improve human-computer interactions, computer vision has been extended to recognize both printed and handwritten characters. Much effort has been carried to recognize English handwritten characters. However, for Asian languages, this is still quite an open topic. Among Asian languages, Sinhala language characters identification is still an open topic since Sri Lanka is the only country that uses Sinhala as the national language.

Compared to the other widely spoken languages, Sinhala language characters have round complicated forms and similarities represent within characters. The Sinhala language includes 60 basic characters (non-cursive). And modifiers that can be added to the core characters expand the Sinhala character set even more. Therefore, the identification of Sinhala handwritten characters is a challenging task.

Most of the research works carried out in character recognition are using pattern matching and image processing techniques. However, these techniques are unable to respond to variations.

This research discusses a handwritten Sinhala character recognition system built using Convolutional Neural Network (CNN). CNN model extract features automatically in the training process, which makes CNN vary from other traditional approaches of Sinhala handwritten character recognition (Sinhala HCR). This research is an endeavour to set a benchmark for Sinhala HCR using the deep learning method. This proposed architecture contains 12 layers including 6 convolutional layers, 3 max-pooling layers, and 3 fully connected layers following batch normalization in each layer. To calculate the probabilities of the output classes, the Softmax function is used. Cross entropy loss is calculated during the training epochs to calculate the loss error and update the network weight accordingly.

The proposed method can recognize 31 Sinhala handwritten characters (7 cursive characters and 24 non-cursive characters) with 99% of training accuracy and 98% of overall testing accuracy. Therefore, this research attempt was able to train a CNN model to identify Sinhala handwritten characters with a high accuracy rate.

**Key Words** – Image Processing, OCR, Convolutional Neural Network, Sinhala Handwritten Character Recognition, Deep Learning