

An Automated Grammar Checking Approach for Sinhala Handwritten Documents

Ashan Indrajith* and W.A.M Prabuddhi

Department of Information and Communication Technology, Faculty of Technology, University of Ruhuna

*ashanindrajith610@gmail.com

Abstract

Sinhala is the mother tongue of most Sri Lankans. The number of Sinhala speakers worldwide is around 20 million. Although many Sri Lankans speak Sinhala fluently, many don't have a deep understanding of its grammatical rules. Special knowledge of Sinhala grammar is required to solve and correct grammatical errors. There are many unique features of the Sinhala language that make it challenging to use it correctly. It is written in letters, and the sentence structure is different from English (subject-object-verb order). Many people find it difficult to apply the grammar rules they learn in the field of education. This leads to common mistakes in written Sinhala. Currently, there are no applications for checking Sinhala grammar in handwritten Sinhala documents. Most existing grammar correction applications require user-typed text and perform the check Sinhala grammar itself only while the user is typing. It can be recognized that there is a significant gap in the resources available for Sinhala to meet the unique challenges posed by handwritten content in particular. Recognizing this issue, this research aims to develop a mobile application that can effectively detect grammar issues in handwritten documents. The proposed solution uses optical character recognition (OCR) technology developed by convolutional neural networks (CNN) to accurately recognize Sinhala characters in handwritten documents. Also, the Natural Language Toolkit (NLTK) and rule-based approach have been used to check the grammatical errors in the identified Sinhala text. The model is trained using a dataset of 7, 0000 handwritten samples collected from various sources covering all Sinhala characters. Apart from this, it is also possible to identify the mistakes in the basic grammar of the Sinhala language by using the rule-based approach. Using these technologies, this application enables all its users to improve their writing skills and develop better language skills, even those who have slight knowledge of grammar rules.

Keywords: *Convolutional Neural Networks, Natural Language Tool Kit, Optical Character Recognition, Rule-based approach, Sinhala handwriting*

**Not presented at the conference*