Handwritten Sinhala Character Recognition Using Deep Learning

M. L. Karunarathne

198762H

Faculty of Information Technology
University of Moratuwa

July 2022

Handwritten Sinhala Character Recognition Using Deep Learning

M. L. Karunarathne

198762H

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master of Science in Information Technology.

July 2022

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student	Signature of Student
M. L. Karunarathne	UOM Verified Signature ………………………………..
	Date:04.08.2022

Supervised by:

Name of Supervisor	Signature of Supervisor
Dr. Chaman Wijesiriwardane	………………………………..
	Date:

Dedication

This effort is dedicated to my family in appreciation of their support, unconditional love, and sacrifice.

Acknowledgements

I would like to acknowledge my supervisor Dr. C. P. Wijesiriwardana for his guidance, encouragement and coordination during the whole research and writing phase. I also like to express my gratitude to Dr. S. C. Premaratne and Dr. W. G. C. W. Kumara for their guidance and suggestions during the research. In addition, I should be grateful to Mr. D. H. M. Jayarathna, a friend of mine, for his support during the research project. A special thanks should go to my family for their unwavering support and understanding when undertaking not only this research but the entire course.

Abstract

Sinhala language is the national language in Sri Lanka. Sinhala alphabet includes 60 characters and is slightly complex compared to other languages like English. Around 25-30 researches have been done since 1990 regarding Sinhala handwritten character recognition. Handwritten Sinhala character recognition remains mostly unsolved in pattern recognition, due to many perplexing characters and excessive curves in Sinhala handwriting. The existing recognizers are also unable to provide acceptable performance for practical applications.

This research aims to enhance the performance of handwritten Sinhala character recognition by using a new approach focused on deep neural networks, which have recently given excellent performance in many applications. This research implements Convolutional Neural Networks (CNNs) and Gabor initialized Convolutional Neural Network (GCNN). In addition to that, it investigates the performance of the proposed network architectures when introducing the dropout. To apply Gabor initialized CNN, the effect of the parameters of the Gabor filter over the Sinhala character image dataset is also examined. Considering the effect of the parameter on the GCNN architecture, parameter values for the proposed GCNN architecture are determined. The training accuracy of the first CNN method is 96.33 % and the testing accuracy is 90.14%. According to the literature, this is the highest accuracy obtained for 60 Sinhala characters compared with primitive methods. This accuracy is obtained with the 0.5 dropout effect. The Gabor initialized CNN architecture provides 95.15% training and 80% testing accuracy. Even though the training accuracy is approximately 1% less than the training accuracy of the first CNN architecture, it converges to the results rapidly. So, it saves time and computational cost.

Considering the results of implemented CNN architectures and Gabor initialized CNN architecture, the best-performing architecture is selected for the Sinhala handwritten character recognition process.

Contents

List of Figures

List of Tables