Universiti Teknologi Mara

Chinese Character Recognition System Using Convolutional Neural Network (CNN)

Nur Fatihah Hatini Binti Mat Sam

Thesis submitted in fulfilment of the requirements
for Bachelor of Computer Science (Hons.) College
of Computing, Informatics and Mathematics

FEBRUARY 2024

ACKNOWLEDGMENT

In the name of Allah, the most Gracious and the most Merciful.

First and foremost, I want to express my sincere gratitude to the Almighty, whose blessings, guidance, and unwavering support have served as a starting point for this research. Through His grace, I have been able to persevere through difficulties and find inspiration throughout my research. I would like to express my sincere gratitude to my supervisor, Ts. Dr. Rajeswari A/P Raju, for her mentorship, patience, and essential advice. Her knowledge, constructive feedback, and support were crucial in determining the direction and quality of this proposal.

Additionally, I would like to express my sincere gratitude to Madam Ummu Fatihah Binti Mohd Bahrin, my lecturer for this course on project formulation. Her perceptive lectures, stimulating discussions, and enthusiasm for the material have all played a significant role in deepening my understanding and forming the conceptual framework of this proposal. Next, I will always be grateful to my parents for their unending support, sacrifices, and love. I've been motivated to learn new things by their unceasing support, faith in my abilities, and willingness to make sacrifices.

Last but not least, I would like to convey my sincere gratitude to my friends, whose support, company, and presence have made my research for this proposal journey fun and memorable. Being able to persevere and conquer obstacles has been made possible by their friendship, compassion, and constant support.

ABSTRACT

This study aims to develop a Chinese character recognition system using Convolutional Neural Network (CNN) system that is specifically designed for the challenging task of recognizing Chinese characters, taking into consideration the complex structures and various styles of writing that these characters possess. With the Chinese language's global importance, particularly in places like Malaysia with one of the large Chinese communities, there is a strong need for an effective and user-friendly character recognition system. The CNN system is remarkably accurate, with a recognition rate of 99.99% for Chinese character recognition and a training accuracy of 92%. The research includes an in-depth examination of CNN requirements, the methodical building of a recognition system, and an accurate evaluation of its performance using testing measures such as precision, recall, and F1-score. The results indicate impressive testing accuracy with minimal loss, demonstrating the system's ability to generalize to newly collected data. The confusion matrix provides an in-depth analysis of the model's classification accuracy over 135 different Chinese character classes. Functionality testing verifies the system's ability to accommodate different handwriting styles, increasing its value for learners of varying competence levels. Recognizing limitations due to the intentional focus on recognizable characters for beginners, the study proposes future directions, such as investigating advanced CNN architectures, augmenting training datasets, incorporating transfer learning strategies, and incorporating user feedback for continuous refinement. In short, this initiative aims to bridge the gap between technological breakthroughs in CNNs and evolving educational objectives for Chinese character recognition, therefore contributing to the larger conversation about the connection of technology and education in linguistic domains.

TABLE OF CONTENT

CONTENT   **PAGE**

CHAPTER 1   INTRODUCTION

CHAPTER 2   LITERATURE REVIEW

CHAPTER 3   METHODOLOGY