

**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 Fatimah 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

<b>CONTENT</b>	<b>PAGE</b>
<b>SUPERVISOR APPROVAL</b>	<b>ii</b>
<b>STUDENT DECLARATION</b>	<b>iii</b>
<b>ACKNOWLEDGMENT</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>TABLE OF CONTENT</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
<b>CHAPTER 1 INTRODUCTION</b>	
1.1 Overview of the chapter	14
1.2 Background of Study	14
1.3 Problem Statement	15
1.4 Objectives	16
1.5 Project Scope	16
1.6 Project Significance	17
1.7 Overview Of Research Framework	18
1.8 Conclusion	19
<b>CHAPTER 2 LITERATURE REVIEW</b>	
2.1 Overview of the chapter	20
2.2 Artificial Intelligence	21
2.3 Pattern Recognition	22
2.3.1 Phase in pattern recognition	23
2.4 Chinese character recognition	24

2.4.1	Chinese character recognition system using Hierarchical Decomposition Embedding	25
2.4.2	Chinese character recognition system using Recurrent Neural Networks (RNN)	27
2.5	Machine Learning	28
2.6	Convolutional Neural Network (CNN)	17
2.6.1	CNN and how does it work?	17
2.6.2	Benefits of CNN	20
2.6.3	Application of CNN algorithm	21
2.7	Implementation CNN algorithm in various problem	25
2.8	Similar works Chinese Character Recognition System	28
2.9	Implication of Literature Review	31
2.10	Conclusion	33

### **CHAPTER 3      METHODOLOGY**

3.1	Overview of the chapter	34
3.2	Research Methodology Framework	35
3.3	Preliminary Study	38
3.3.1	Literature Review	38
3.4	Data collection	40
3.4.1	Data Description	40
3.4.2	Pre-Processing Data	42
3.5	Prototype Design	43
3.5.1	System Architecture	43
3.5.2	System flowchart	46
3.5.3	User Interface Design	48
3.6	Implementation and Development	49
3.6.1	Model Training and Optimizing	49
3.7	Testing and Evaluation	49
3.7.1	Confusion Matrices	50