Universiti Teknologi MARA

Arabic Character Recognition System Using Convolutional Neural Network (CNN)

Nurul Amira Binti Abdullah

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

FEBRUARY 2024

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah S.W.T because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. I would like to express my sincere gratitude and appreciation to all those who have contributed to the completion of this research on Arabic Handwritten Recognition using Convolutional Neural Networks (CNN).

First and foremost, I am grateful to my supervisor Sir Zulkifli Bin Halim for his guidance, support, and valuable insights throughout the research process. His expertise and encouragement have been instrumental in shaping this study.

I extend my heartfelt thanks to my lecturer, Ummu Fatihah Binti Mohd Bahrin for her valuable feedback and suggestions, which have significantly contributed to the improvement of this work.

I would like to acknowledge the participants who generously provided their handwritten Arabic text samples for the dataset used in this research. Their contributions have been invaluable in enabling the development and evaluation of the CNN model.

I am indebted to the researchers and scholars in the field of Arabic Handwritten Recognition using CNN, whose previous work and publications have laid the foundation for this research. Their insights and findings have served as a valuable reference and inspiration throughout this study.

I would also like to express my gratitude to my family and friends for their unwavering support and encouragement. Their belief in my abilities has been a constant source of motivation.

Lastly, I would like to thank the academic institutions, funding organizations, and research facilities that have provided resources and facilities enabling the successful completion of this research endeavor. Without the support and contributions of all those mentioned above, this research would not have been possible. Thank you for your invaluable assistance and guidance throughout this journey.

ABSTRACT

This research project focuses on Arabic Handwritten Recognition system using Convolutional Neural Networks (CNNs) algorithm. This study delves into the challenging realm of Arabic handwriting recognition, spurred by the intricate nature of the script and the scarcity of specialized tools and high-quality training data. The investigation primarily focuses on the effectiveness of Convolutional Neural Networks (CNNs) in mitigating these challenges through the development of a Handwritten Character Recognition System (HCR) tailored for Arabic script. Leveraging CNNs, the system endeavors to accurately transcribe and comprehend handwritten Arabic documents, thereby facilitating efficient processing and analysis. Through a comprehensive literature review, the research underscores the significance of Arabic handwriting recognition across various domains, such as document digitization, archival systems, historical document analysis, and language learning, particularly among toddlers. Methodologically, the study adopts a structured seven-phase approach, commencing with a preliminary study encompassing a comprehensive literature review to identify the project's objectives, scope, and significance. Subsequent phases include requirement analysis, data collection, prototype design, implementation, evaluation, and documentation. By meticulously collecting and partitioning a dataset of 16,800 handwritten Arabic characters into training and test sets, the project rigorously trains and evaluates the CNN model, achieving commendable accuracy rates of 92.50% and 97.35% in testing and training, respectively. Beyond technical advancements, the research emphasizes broader societal implications, such as cultural preservation, educational benefits, and technological advancements. However, identified limitations, notably in handling diverse character variations and computational requirements, underscore the need for continued research and refinement. Overall, this project signifies a significant milestone in the field of Arabic character recognition, promising broader accessibility to Arabic language materials and enriching cultural heritage preservation efforts through transformative technological advancements.

TABLE OF CONTENTS

CONTENT			PAGE
SUPERVISOR APPROVAL			ii
STUDENT DECLARATION			iii
ACKNOWLEDGEMENT			iv
ABSTRACT			v
TABLE OF CONTENTS			vi
LIST OF FIGURES			ix
LIST OF TABLES			xi
LIST OF ABBREVIATION			xii
CHAPTER ONE: INTRODUCTION			
1.1	Background of Study		1
1.2	Problem Statement		3
1.3	Objectives		3
1.4	Project Scope		4
1.5	Significance of the Project		4
1.6	Overview of Research Framework		6
1.7	Conclusion		7
CHAPTER TWO: LITERATURE REVIEW			
2.1	Introduction		8
2.2	Handwriting Recognition System		10
	2.2.1	Overview of Handwriting Recognition System	10
	2.2.2	The Use of Handwriting Recognition System	11
	2.2.3	Arabic Handwriting Recognition	13

2.3 Machine Learning 15
2.3.1 Overview of Machine Learning 15
2.3.2 Machine Learning Technique 16
2.4 Convolutional Neural Network 20
2.4.1 CNN and How Does It Works 20
2.4.2 Benefit of CNN's 22
2.4.3 Implementation of CNN Algorithm in Various Problem 23
2.5 Similar Works 27
2.6 The implication of Literature Review 31
2.7 Conclusion 32

CHAPTER THREE: METHODOLOGY

3.1 Overview of Research Methodology 33
3.1.1 Detailed of Research Framework 33
3.2 Research Phases 37
3.3 Preliminary Study 38
3.4 Knowledge Acquisition 39
3.5 Data Collection 40
3.5.1 Data Description 41
3.5.2 Data Pre-processing 42
3.6 Design Phase 43
3.6.1 System Architecture (Logical Design) 44
3.6.2 Flowchart (System Development Flowchart) 45
3.6.3 User Interface Design 47
3.7 Prototype Implementation 48
3.8 Performance Evaluation 49
3.8.1 Confusion Matrices 50
3.9 Gantt Chart 53
4.0 Conclusion 54