



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

**A STUDY OF HANDWRITTEN TEXT CHARACTER
RECOGNITION USING NEURAL NETWORK**

This report is submitted in accordance with the requirement of the Universiti Teknikal Malaysia Melaka (UTeM) for the Bachelor of Electronic Engineering Technology (Telecommunications) with Honours.



NUR MAISARAH BINTI HASSAN

B071710582

961031-56-5120

FACULTY OF ELECTRICAL AND ELECTRONIC ENGINEERING
TECHNOLOGY

2020

BORANG PENGESAHAN STATUS LAPORAN PROJEK SARJANA MUDA

Tajuk: **A STUDY OF HANDWRITTEN TEXT CHARACTER RECOGNITION USING NEURAL NETWORK**

Sesi Pengajian: **2019 / 2020**

Saya **NUR MAISARAH BINTI HASSAN** mengaku membenarkan Laporan PSM ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka (UTeM) dengan syarat-syarat kegunaan seperti berikut:

1. Laporan PSM adalah hak milik Universiti Teknikal Malaysia Melaka dan penulis.
2. Perpustakaan Universiti Teknikal Malaysia Melaka dibenarkan membuat salinan untuk tujuan pengajian sahaja dengan izin penulis.
3. Perpustakaan dibenarkan membuat salinan laporan PSM ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ****Sila tandakan (X)**

- SULIT** Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia sebagaimana yang termaktub dalam AKTA RAHSIA RASMI 1972
- TERHAD** Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan
- TIDAK TERHAD**

Yang Benar,

Disahkan oleh :

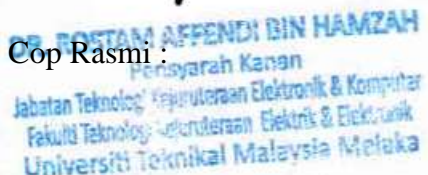


(TANDATANGAN PENULIS)



(TANDATANGAN PENYELIA)

Alamat Tetap :
NO. 15, JALAN PRIMA
SAUJANA 2/3, TAMAN PRIMA
SAUJANA SEKSYEN 2, 43000,
KAJANG, SELANGOR

Cop Rasmi :

DR. ROSLAN AFFENDI BIN HAMZAH
Penyarah Kanan
Jabatan Teknologi Kejuruteraan Elektronik & Komputer
Fakulti Teknologi Kejuruteraan Elektrik & Elektronik
Universiti Teknikal Malaysia Melaka

Tarikh : 29/1/2021

Tarikh : 29/1/2021

*Jika Laporan PSM ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan PSM ini

DECLARATION

I hereby, declared this report entitled A STUDY OF HANDWRITTEN TEXT CHARACTER RECOGNITION USING NEURAL NETWORK is the results of my own research except as cited in references.



Signature :
Author : NUR MAISARAH BINTI HASSAN
Date : 29 JANUARY 2021



اونيورسيتي تيكنيكل مليسيا ملاك
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

APPROVAL

This report is submitted to the Faculty of Electrical and Electronic Engineering Technology of Universiti Teknikal Malaysia Melaka (UTeM) as a partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering Technology (Telecommunication) with Honours. The member of the supervisory is as follow:



ABSTRAK

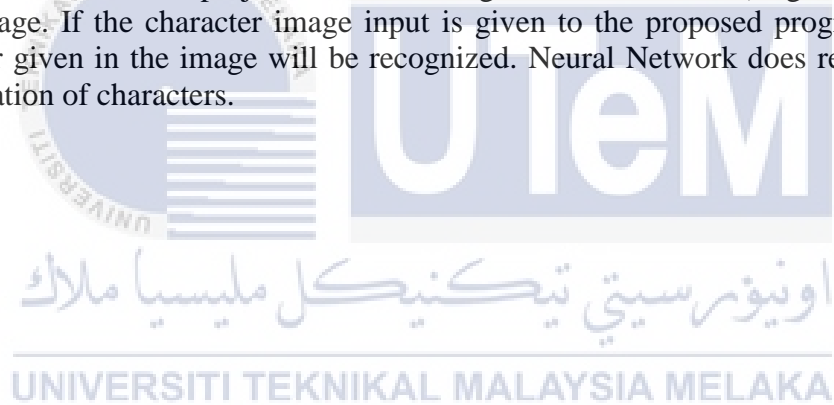
Projek tajuk ini mencadangkan pendekatan yang efisien terhadap pengembangan sistem pengenalan teks tulisan tangan menggunakan Neural Network. Oleh kerana ketidakkonsistenan dalam gaya tulisan tangan, sistem pengenalan tulisan tangan yang canggih sering kali gagal memberikan prestasi yang memuaskan pada pelbagai jenis sampel tulisan tangan. Pendekatan yang tersedia untuk pengenalan tulisan tangan biasanya terdiri daripada pelbagai langkah yang merangkumi terutamanya 1.Pre-pemprosesan, 2.Pengekstrakan ciri, 3.Pengelasan dan 4. Pemprosesan pasca. Objektif utama kajian ini adalah untuk mengembangkan sistem pengenalan watak dan nombor tulisan tangan yang cekap untuk watak Inggeris berdasarkan Artificial Neural Network. Aksara tulisan tangan mungkin mengandungi huruf besar (huruf besar dan huruf kecil) watak Inggeris dengan 52 kelas (26 untuk huruf besar dan 26 untuk huruf kecil) termasuk dalam kajian ini untuk klasifikasi. Perisian pengiraan MATLAB dengan Image Processing Toolbox dan Neural Network Toolbox akan digunakan untuk menyelesaikan masalah pengelasan watak tulisan tangan yang ditentukan dari klasifikasi. Projek ini berfungsi untuk mengenali semua watak (Bahasa Inggeris) yang disediakan sebagai gambar input. Sekiranya input imej watak diberikan kepada program yang dicadangkan, watak input yang diberikan dalam gambar akan dikenali. Neural Network melakukan pengelasan dan pengelasan watak.

اونيورسي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

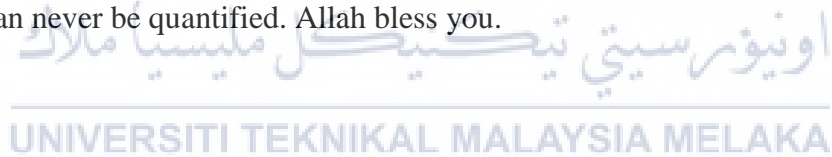
ABSTRACT

This title project uses Neural Network as a method to develop recognition systems for handwritten text characters. The handwriting recognition systems frequently fail or unable to give sufficient results on different types of handwriting due to massive inconsistency styles of handwriting. Handwriting recognition systems comprise of Pre-processing, Segmentation, Feature Extraction and Classification. The main goal of this project is to propose a framework of text character recognition algorithm using Neural Network. The classification comprises 52 classes of English handwritten characters with 26 characters for capital letters and 26 characters for small letters. The handwritten character recognition will be using MATLAB software with Image Processing and Neural Network Toolbox. This project serves to recognize all characters (English) provided as input image. If the character image input is given to the proposed program, the input character given in the image will be recognized. Neural Network does recognition and classification of characters.



DEDICATION

I dedicate this project to My Creator Allah S.W.T, my strong pillar, my source of inspiration, wisdom, knowledge and understanding. He has been the source of my strength throughout this project and on His wings only have I soared. I also dedicate this work to families; Hassan bin Mat Lihim, my dad and Zainab binti Ismail, my mom, who has encouraged me all the way and whose encouragement has made sure that I give it all it takes to finish that which I have started. To my sister, Umami Umairah (Adik), and fellow friends who have been affected in every way possible by this quest. This project I also dedicate to my supervisor, Ts. Dr. Rostam Affendi bin Hamzah for his help to manage me to the accomplishment of undertaking for my degree. Thank you. My love for you all can never be quantified. Allah bless you.



ACKNOWLEDGEMENTS

“In the name of Allah S.W.T, the Most Gracious and the Most Merciful”

Alhamdulillah, Praise to Allah S.W.T for his blessing and guidance have helped me carry out my project and thesis completely. First, I would like to express my gratitude to my supervisor, TS. DR. ROSTAM AFFENDI BIN HAMZAH. He gave me support and monitor my progress and my performance on this project. I'm grateful to my supervisor who always being my guidance and his kindness for accepting me as his student final year's project will be always remembered. I have learned a lot while under his guidance in practically or theoretically.

My sincere thanks I dedicate my special appreciation and thanks to my family especially my beloved parent HASSAN BIN MAT LIHIM and ZAINAB BINTI ISMAIL and to all my family members who always give me support and encourage me to give my best. Their full support and encouragement were always give me strength to keep continuing doing this project.

Lastly, I would like to thanks to all my lecture and my friend for their invaluable assistance toward this project. Without support from them, it was impossible for me to complete this project.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

TABLE OF CONTENTS

	PAGE
TABLE OF CONTENTS	x
LIST OF TABLES	xiv
LIST OF FIGURES	xvi
LIST OF GRAPHS	xviii
LIST OF APPENDICES	xix
LIST OF SYMBOLS	xx
LIST OF ABBREVIATIONS	xxi
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Background	1
1.3 Problem Statement	4
1.4 Objective	5
1.5 Scope	5
1.6 Thesis Outline	6
CHAPTER 2 LITERATURE REVIEW	7
2.1 Introduction	7
2.2 Relevent Theory	7
2.2.1 Overview of Handwritten Text Character Recognition	7

2.2.2	Neural Network	12
2.3	Methods or Techniques Approach in Handwritten Text Character Recognition	17
2.3.1	Research on a Handwritten Recognition Algorithm based on an Extended Nonlinear Kernel Residual Network	17
2.3.2	Handwritten Gurmukhi Character Recognition	18
2.3.3	Use of Gradient Technique for Extracting Features from Handwritten Gurmukhi Character and Numerals	19
2.3.4	Neural Network based Handwritten Character Recognition System	20
2.3.5	Handwritten Text Recognition System based on Neural Network	20
2.3.6	Online and Offline Handwritten Chinese Character Recognition : A Comprehensive Study and New Benchmark	21
2.3.7	Neural Architecture Based on Fuzzy Perceptual Representation for Online Multilingual Handwriting Recognition	23
2.3.8	Handwritten Character Recognition Using Unique Feature Extraction Technique	23
2.3.9	Offline Handwritten Hindi 'SWARs' Recognition Using a Novel Wave Based Feature Extraction Method	24
2.3.10	Recognition of Characters in Indian MODI Script	24
2.3.11	Sindhi Handwritten-Digits Recognition Using Machine Learning Techniques	24
2.3.12	Online Farsi Handwritten Character Recognition Using Hidden Markov Model	25

2.3.13	Fuzzy-Based Multi-Kernel Spherical Support Vector Machine for Effective Handwritten Character Recognition	25
2.3.14	Online Bangla Handwritten Word Recognition Using HMM and Language Model	26
2.3.15	Handwritten Arabic Numeral Character Recognition Using Multi Kernel Learning Support Vector Machine	27
2.4	Comparison Between Difference Methods or Techniques Approach in Handwritten Text Character Recognition	29
2.5	Technologies in Handwritten Text Character Recognition	39
2.5.1	Pen to Print	39
2.5.2	Prime Scanner	40
2.5.3	SimpleOCR	43
CHAPTER 3 METHODLOGY اوينورسيتي تيكنيكل مليسيا ماراڤ		46
3.1	Introduction	46
3.2	The Project Execution Flowchart	47
3.3	Project Implementation	49
3.3.1	Project Planning	49
3.4	Flowchart Represent Process of Project	52
3.5	Proposed Method	53
3.5.1	Image Acquisition	54
3.5.2	Pre-processing	54

3.5.3	Segmentation	56
3.5.4	Feature Extraction	57
3.5.5	Classification	58
3.6	Software Implementation	61
3.6.1	MATLAB Software	61
CHAPTER 4 RESULT AND DISCUSSION		64
4.1	Introduction	64
4.2	Software Simulation	64
4.2.1	Coding Process of The System	64
4.3	Experimental Result	70
4.3.1	26 Upper and Lower Case Letters with Different Handwritings	72
4.3.2	10 Different Handwritings with 10 Words	78
4.3.3	Analysis Results	81
CHAPTER 5 CONCLUSION AND RECOMMENDATION		90
5.1	Introduction	90
5.2	Conclusion	90
5.3	Recommendation	91
REFERENCES		92
APPENDIX		96

LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1:	Comparison of Offline and Online Handwriting Recognition	11
Table 2.2:	Advantages of Neural Network	16
Table 2.3:	Comparison Between Difference Methods or Techniques Approach	30
Table 2.4:	Features of SimpleOCR	44
Table 3.1:	Project Planning	49
Table 3.2:	Project Gantt Chart PSM 1	50
Table 3.3:	Project Gantt Chart PSM 2	51
Table 3.4:	Neural Network Feed Forward Back Propagation Training Parameters	59
Table 3.5:	Key Features of MATLAB Image Processing Toolbox	62
Table 3.6:	Key Features of MATLAB Neural Network Toolbox	63
Table 4.1:	26 Upper Case Letters with Different Handwritings	72
Table 4.2:	26 Lower Case Letters with Different Handwritings	75
Table 4.3:	10 Different Handwritings with 10 Words (UPPER CASE)	78
Table 4.4:	10 Different Handwritings with 10 Words (LOWER CASE)	79
Table 4.5:	10 Different Handwritings with 10 Words (MIX)	80
Table 4.6:	Results for Upper and Lower Case Letters with Different Handwritings	81
Table 4.7:	Results for 10 Different Handwritings With 10 Words (UPPER CASE)	82

Table 4.8: Results for 10 Different Handwritings with 10 Words
(LOWER CASE) 83

Table 4.9: Results for 10 Different Handwritings with 10 Words (MIX) 84



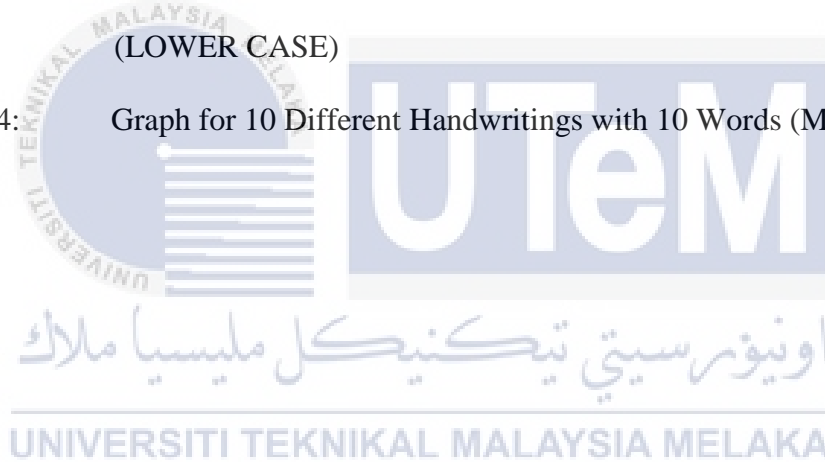
LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 1.1:	Generic Character Recognition Model	3
Figure 2.1:	Handwritten Text Character Recognition	7
Figure 2.2:	The Comparison of Offline and Online Handwriting Recognition	11
Figure 2.3:	A Simple Neural Network	12
Figure 2.4:	Neuron	14
Figure 2.5:	Extended Nonlinear Kernel Residual Network Framework	18
Figure 2.6:	Samples of Handwritten Gurmukhi	19
Figure 2.7:	Block diagram of the testing of ANN	21
Figure 2.8:	directMap Framework	22
Figure 2.9:	convNet Architecture	22
Figure 2.10:	Handwritten Character Recognition using Fuzzy-Based Multi-Kernel Spherical Support Vector Machine Block Diagram	26
Figure 2.11:	Multi Kernel Support Vector Machine	27
Figure 2.12:	Pen to Print Logo	39
Figure 2.13:	Handwriting to Text App	39
Figure 2.14:	Prime Scanner Logo	40
Figure 2.15:	OCR Scanner without Internet	41
Figure 2.16:	Prime Scanner Application with OCR	41
Figure 2.17:	SimpleOCR	43

Figure 3.1:	Planning Project Flowchart	48
Figure 3.2:	Flowchart Represents Process of Project	52
Figure 3.3:	Block Diagram of the Proposed Recognition System	54
Figure 3.4:	The Process of Pre-Processing Method in Handwritten Character Recognition	56
Figure 3.5:	Original Image to Normalization Image	57
Figure 3.6:	The Process of Feature Extraction From The Characters	58
Figure 3.7:	Neural Network Feed Forward Back Propagation Architecture	59
Figure 3.8:	The Output will be The Beautified Version of The Uploaded Image	60
Figure 3.9:	MATLAB Software	61
Figure 4.1:	Code for Image Pre-processing	65
Figure 4.2:	Input Image	65
Figure 4.3:	Grayscale Data Type	66
Figure 4.4:	Binary Image	67
Figure 4.5:	Binary Image Data Type	67
Figure 4.6:	Threshold Optimum Value	67
Figure 4.7:	Code For Segmentation	68
Figure 4.8:	Segmented Image	69
Figure 4.9:	Code for Feeding Neural Network and Detecting Text	70
Figure 4.10:	E-MNIST Dataset	71

LIST OF GRAPHS

GRAPH	TITLE	PAGE
Graph 4.1:	Graph of 26 Upper and Lower Case Letters with Different Handwritings	85
Graph 4.2:	Graph for 10 Different Handwritings with 10 Words (UPPER CASE)	86
Graph 4.3:	Graph for 10 Different Handwritings with 10 Words (LOWER CASE)	87
Graph 4.4:	Graph for 10 Different Handwritings with 10 Words (MIX)	88



LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix 1:	Coding for Handwritten Character Recognition	96
Appendix 2:	Coding for Image Labeller	98
Appendix 3:	Coding for Training the Neural Network	99



LIST OF SYMBOLS

$g(x,y)$	-	Grayscale Image
$R(x,y)$	-	Red Component of the Colour Image
$G(x,y)$	-	Green Component of the Colour Image
$B(x,y)$	-	Blue Component of the Colour Image
w^i	-	Weight Vector of i^{th} Layer
a^i	-	Output of i^{th} Layer
b^i	-	Bias Vector for i^{th} Layer



LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
ACSII	-	American Standard Code for Information Interchange
OCR	-	Optical Character Recognition
ICR	-	Intelligent Character Recognition
RNN	-	Recurrent Neural Network
PDA	-	Personal Digital Assistant
LCD	-	Liquid Crystal Display
SVM	-	Support Vector Machine
MLP	-	Multi Layered Perceptron
CPU	-	Central Processing Unit
MNIST	-	Modified National Institute of Standards and Technology Dataset
SVHN	-	Street View House Numbers Dataset
SVM-RBF	-	Support Vector Machine Radial Basis Function
HCCR	-	Handwritten Chinese Character Recognition
ICDAR	-	International Conference on Document Analysis and Recognition
CNN	-	Convolutional Neural Network
LSTM	-	Long Short Term Memory Network
MLP-BP	-	Multilayer Perceptron Network using Backpropagation Algorithm

MLP-LM	-	Multilayer Perceptron Network using Lavenberg-Marquadt Algorithm
MDC	-	Minimum Distance Classifier
ANESP	-	Automated Numeral Extraction And Segmentation Program
k-NN	-	k-Nearest Neighbour
ML	-	Machine Learning
DT	-	Decision Tree
HMM	-	Hidden Markov Model
LM	-	Language Model
HOG	-	Histogram of the Oriented Gradient
MKL	-	Multi Kernal Learning
HTML	-	Hypertext Markup Language
TIFF	-	Tagged Image File Format
TXT	-	Text File
RTF	-	Rich Text Format
PDF	-	Portable Document Format
ROI	-	Region Of Interest
ICC	-	International Color Consortium Profile
DICOM	-	Digital Imaging And Communication in Medicine
LVQ	-	Learning Vector Quantization
NARX	-	Nonlinear Autoagressive
GPU	-	Graphics Processing Unit

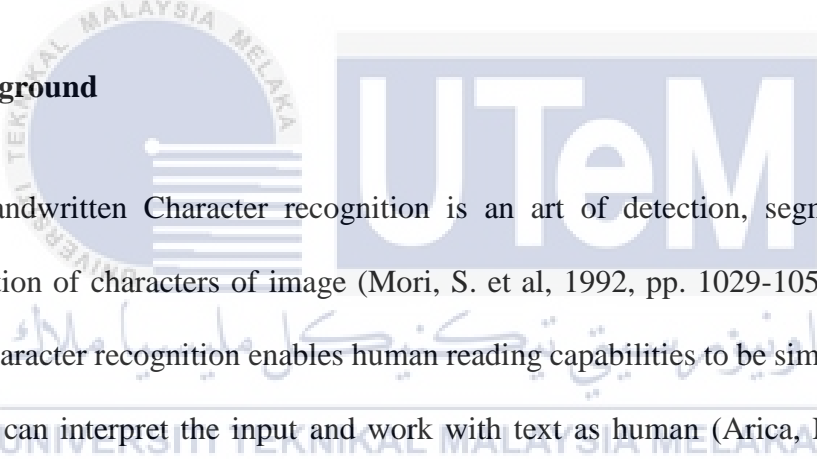
CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter will further unfold a recap of this project regarding Handwritten Text Character Recognition using Neural Network. This chapter will review about background, statement of problem, objective, scope, and project significance. The thesis outline is mentioned at the end of this section.

1.2 Background



Handwritten Character recognition is an art of detection, segmentation, and identification of characters of image (Mori, S. et al, 1992, pp. 1029-1058). This handwritten character recognition enables human reading capabilities to be simulated thus the computer can interpret the input and work with text as human (Arica, N. & Yarman-Vural, F.T., 2001, pp. 216-233). According to (Mohapatra, H., 2009), the identification of handwriting has been one of the most important and difficult groundwork areas in recent years in the field of image processing and pattern recognition. It makes an important contribution to the automation process development, and in many applications strengthens the interface between man and machine. The main goal of this project is to propose a framework of text character recognition algorithm using Neural Network. To put it another way, a recognition using the Artificial Neural Network (ANN) approach can functionally identify a type of particular character.

Neural computing is a comparatively new area and therefore the design components are barely well defined compared to other frameworks. Data Parallelism applies to neural computers. Neural computers operate in a manner entirely different from standard computer performance. Neural computers are not programmed but trained to provide a certain starting state of data input, either classify the input data into one of the class numbers or cause the original data to evolve in a way that enhances some preferable property.

Recognition of characters consists of two types; online and offline. In online recognition, data is gathered in online character recognition during the writing process with the aid of a special pen on the electronic surface. Pre-written data written on a sheet of paper is scanned for offline recognition. Online approaches have been proven to be an expert on the identification of handwritten characters by their off-line counterparts, due to the temporal knowledge available to the former. According to (Bhattacharya, U. & Chaudhuri, B.B., 2008, pp. 444-457), the level of accuracy in recognition in off-line systems is comparably high. The handwritten characters include for classification in this analysis is a mix case (capital and small letters) of English characters consisting of 52 groups (26 for capital letters and 26 for smaller letters).

Based on Figure 1.1, recognition of characters consists of two sections; that is, section of training and test. The handwritten character images are scanned in both the training and testing section, followed by pre-processing process which includes noise filtering, smoothing, and normalization. Image rendering is ideal for segmentation in which the image is decomposed into sub-images. Meanwhile, Feature Extraction functions in improving identification rate and misclassification. The finishing process of the reconnaissance system is the simulation of the trained classifier. It produces

recognized representative characters by measuring equal of standardized ASCII character type, which means the Test Sample Recognition Index will be used.

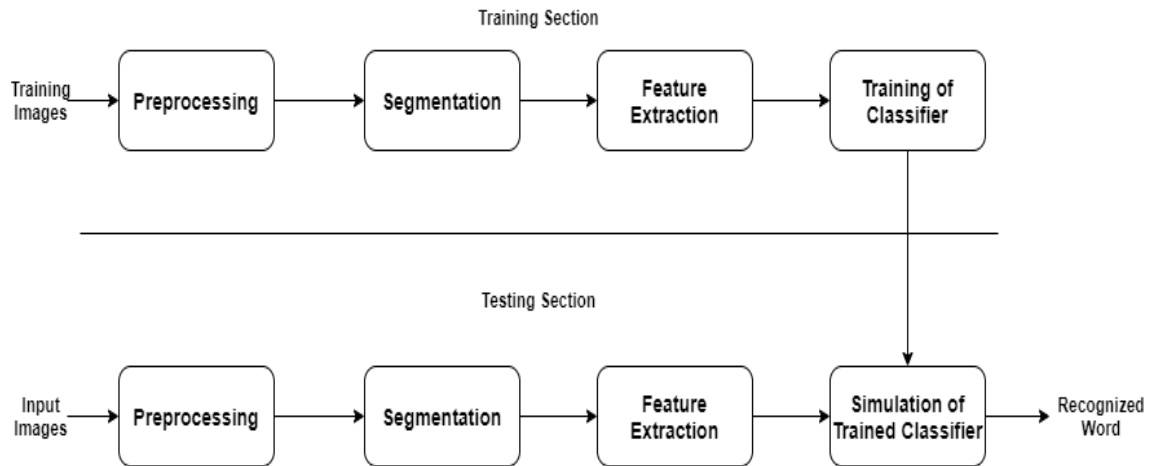


Figure 1.1 : Generic Character Recognition Model

Recently, electronic handwritten recognition has drawn the attention of researchers worldwide due to the increased use of handheld devices. Nowadays, many applications have included offline handwriting recognition systems, including mail sorting, bank processing, document reading and recognition of postal addresses. Consequently, the recognition of off-line handwriting continue to be an active area of research to explore the latest techniques that would improve recognition accuracy (Pal, U. et al, 2007, pp. 749-753) (Pal, U. & Chaudhuri, B.B., 2004, pp. 1887-1899).