Novel Feature Extraction Methods to Automatically Detect Gender and Age From Handwritten Documents

Najla Saleh AL-Qawasmeh

A Thesis

in

The Department

of

Computer Science and Software Engineering

Center of Pattern Recognition and Machine Intelligence CENPARMI

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Computer Science and Software Engineering) at

Concordia University

Montréal, Québec, Canada

September 2022

© Najla Saleh AL-Qawasmeh, 2022

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mrs. Najla Saleh AL-Qawasmeh**

Entitled: **Novel Feature Extraction Methods to Automatically Detect Gender and Age From Handwritten Documents**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science and Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Anjan Bhowmick

_____ External Examiner
Dr. Laurence Likforman

_____ Examiner
Dr. David Morris

_____ Examiner
Dr. Jinqiu Yang

_____ Examiner
Dr. Olga Ormandjieva

_____ Supervisor
Dr. Ching Y. Suen

Approved by _____
Lata Narayanan, Chair
Department of Computer Science and Software Engineering
Center of Pattern Recognition and Machine Intelligence CENPARMI

September 13, 2022

_____
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Novel Feature Extraction Methods to Automatically Detect Gender and Age From Handwritten Documents

Najla Saleh AL-Qawasmeh, Ph.D.

Concordia University, 2022

Handwriting acts as a mirror, reflecting the writer's personality characteristics and demographic properties. As a result, handwriting analysis has become significant in a variety of disciplines, including medical, socialisation, and security. Identifying a person's biometric information from handwriting has recently become a crucial research topic. Therefore, we decided to focus our research on extracting gender and age from handwriting, as the former is important in psychology, document analysis, paleography, graphology, and forensic investigation. The latter on the other hand, is essential in medical diagnosis and forensic analysis.

Handwriting analysis is commonly done by extracting relevant features such as slant, pen pressure, word spacing, and others. The traditional method of studying handwritten texts involved a graphologist visually inspecting a set of related features. However, analysing many handwritings takes a long time, making it a tiring and time-consuming process. In addition, the analyst's level of expertise influences the analysis too. As a result, we have studied and developed an automatic technique for analysing handwriting documents to extract the writer's gender and age without additional human intervention.

The proposed analysis systems include the five essential aspects of handwriting analysis systems. An Arabic dataset was first gathered to test and evaluate the proposed systems. We named it Free Style Handwritten Samples (FSHS). Our dataset stands out from others because it has 2200 writers, more than any other Arabic dataset currently available. This made the dataset vary with the

handwriting styles, as the total number of handwritten documents in the dataset is 2700. The dataset is mostly text-independent, except 500 samples are text-dependent. The second phase is image acquisition and preprocessing. Moreover, to extract the related features, image processing techniques and the transfer learning method were utilised. While in the classification phase, Support Vector Machine (SVM) and Neural Network (NN) algorithms were employed. Finally, multiple assessment metrics were used to evaluate the proposed systems.

Automatically identifying a writer's gender and age is a challenging task due to the overlapping features of handwritings from different writers. Auto-gender and age detection can be implemented using a set of extracted features or by applying transfer learning. This research presents two different methodologies to detect the gender and age of a writer from handwritten documents. First, a set of related features were extracted for the gender detection system, such as pen pressure, word spacing, text line irregularity, and left and right margins. At the same time, the percentage of black and white pixels and irregularities in pen pressure, slant, and text lines were retrieved to determine the age of writer. Then, the extracted features were analysed separately, followed by the generation and testing of combinations of two and three features. Finally, the sum of all features was evaluated. Subsequently, an automatic feature extraction methodology leveraging the knowledge from two Convolutional Neural Networks (CNN), GoogleNet and ResNet, was applied.

The two proposed systems performed well when applied to the FSHS dataset. The accuracy rates for the gender detection system employing SVM and NN methods, respectively, were 94.7% and 97.1%. In addition, when we used the SVM and NN approaches to apply the age detection system to the same dataset, we obtained accuracy values of 71% and 63.5%, respectively.

To compare our results to other available works, the proposed systems were applied to the public datasets ICDAR2013 and the Khatt dataset. Finally, a number of assessment and tests were carried out to assess the effectiveness of the proposed systems, and the results reveal that the proposed gender and age detection systems outperform the current state-of-the-art with an accuracy rates of 67% and 96.2%, respectively.

Acknowledgments

I would like to gratefully offer my gratitude and thanks to my supervisor Dr. Ching Y. Suen. It has been my good luck, to be supervised by him, because of his invaluable advice, creative vision and support. I also appreciate the financial support from Gina Cody School of Engineering and Computer Science and NSERC (Natural Sciences and Engineering Research Council of Canada.

I am also thankful to my parents, my beloved mother and father, for their continued support and advises throughout my studying years which helped me to achieve my dreams so far. Their presence in my life has enlightened my way to success. They gave me my name, they gave me my life, and everything else in between. I can't find the words that can express how much I love them and appreciate the efforts they have put into giving me the life I have now. They are the reason, I thrive to be better. Their pride of me is the main goal of my life. This research would not be possible without the support of Allah and them.

I want to convey my sincere gratitude to, my daughter Cady, who is my greatest love in this life, for bringing me such endless happiness and pleasure. She enabled me to appreciate life's beauty and to unwind when things became tough.

Brothers and sisters, thank you for believing in me. You are all participants in this achievement.

Last but not least, thanks to everyone who helped in the completion of this thesis, whether directly or in directly.

Contents

List of Figures viii

List of Tables xiii

1 Introduction 1

1.1 Overview 1

1.2 Arabic Language 5

1.3 Motivations 8

1.4 Objectives 9

1.5 Challenges 11

1.6 Contributions 11

1.7 Thesis Outlines 12

2 Literature Review 13

2.1 Datasets 13

2.2 Gender Detection 18

2.3 Age Detection 26

3 Proposed Framework 31

3.1 Dataset Preparation and Image Acquisition 31

3.2 Preprocessing 38

3.3 Feature Extraction 40

3.3.1 Feature Extraction Using Image Processing Techniques 40

3.3.2 Age Related Features 51
3.4 Classification 62

4 Experiments and Results **66**
4.0.1 Experimental Setup 66
4.0.2 Evaluation 68

5 Conclusion and Future Works **99**
5.1 List of Publications 102

Appendix A Appendix A **103**

Appendix B My Appendix **106**

Bibliography **110**

List of Figures

Figure 1.1 General structure of computerized handwriting analysis system 4
Figure 1.2 Different fonts of Arabic language 5
Figure 1.3 Characters with the same shape but different dot positions 6
Figure 1.4 Diacritics in Arabic language 6
Figure 1.5 Ligatures in Arabic language 7
Figure 1.6 Connectivity in Arabic language 7
Figure 1.7 Different letter shapes in Arabic language 7
Figure 1.8 Main processes of the proposed system 10
Figure 3.1 General structure of the proposed systems 32
Figure 3.2 Samples of the dataset (FSHS) in terms of gender and age 33
Figure 3.3 Information page of samples in the dataset 34
Figure 3.4 Text dependent samples of FSHS dataset 34
Figure 3.5 Example of a discarded handwritten document from the FSHS dataset . . . 35
Figure 3.6 Comparison between our dataset and other datasets 36
Figure 3.7 Distribution of writers with respect to (a) Occupation, (b) Age, (c) Gender and (d) Handedness 37
Figure 3.8 Preeprocessing in terms of binarization and filtering 39
Figure 3.9 The cropping of handwriting 39
Figure 3.10 Example of dilated handwriting document 40
Figure 3.11 Difference between female and male pen pressure 42
Figure 3.12 SD Distribution of a set of handwriting samples 42

Figure 3.13 Arabic handwritten text-line 42
Figure 3.14 Diacritics removing process in text segmentation 44
Figure 3.15 Comparison between different accuracy rates of different word spacing bins 45
Figure 3.16 The Difference between female and male handwriting in terms of word spaces 45
Figure 3.17 Margin difference between female and male handwriting 46
Figure 3.18 Right margin feature in female and male handwriting 47
Figure 3.19 Left margin feature in female and male handwriting 47
Figure 3.20 Comparison of different accuracy rates for different margin lengths 48
Figure 3.21 Overlapping between text-lines 49
Figure 3.22 Accuracy rates of different projection profiles where (a) vertical projection and (b) horizontal projection 50
Figure 3.23 Different combinations of horizontal and vertical projection profiles 50
Figure 3.24 Vertical projection profile of two different documents 51
Figure 3.25 Horizontal projection profile of two different documents 51
Figure 3.26 Corresponding accuracy rates using different square numbers 53
Figure 3.27 Corresponding accuracy rates using different square numbers 53
Figure 3.28 Corresponding accuracy rate of different number of CCs 54
Figure 3.29 Handwriting samples with their corresponding slant irregularity distribution 54
Figure 3.30 Corresponding accuracy rates using different square numbers 55
Figure 3.31 Corresponding accuracy rates using different square numbers 56
Figure 3.32 Corresponding accuracy rates using different square numbers 56
Figure 3.33 Handwriting samples with their corresponding pen pressure irregularity distribution 57
Figure 3.34 Corresponding accuracy rates using different square dimensions 58
Figure 3.35 Corresponding accuracy rates using different square dimensions 58
Figure 3.36 Corresponding accuracy rates using different square dimensions 58
Figure 3.37 Distribution of the number of black pixels in youth and mature handwriting 59
Figure 3.38 Traditional machine learning vs transfer learning 60
Figure 3.39 Residual learning: a building block 61

Figure 3.40 ResNet architecture 62
Figure 3.41 Building blocks of GoogleNet 63
Figure 3.42 Structure of transfer learning method 63
Figure 3.43 Basic Neural Network architecture 65
Figure 4.1 Handwritten samples from ICDAR2013 dataset 67
Figure 4.2 Samples of handwritten documents from FSHS dataset 68
Figure 4.3 Samples of handwritten documents from Khatt dataset 68
Figure 4.4 Comparison of feature importance between SVM and NN 72
Figure 4.5 ROC comparison between SVM and NN results using individual features on (a) Margins, (b) Word spacing, (c) Irregularity, and (d) Pen pressure, all applied on FSHS dataset. 73
Figure 4.6 ROC comparison between SVM and NN results using two features combination on (a) LRM + WS, (b) LRM + TLI, (c) LRM + PP and (d) WS + TLI, (e) WS + PP, and (f) TLI + PP, all applied on FSHS dataset. 74
Figure 4.7 ROC comparison between SVM and NN results using three features combinations on (a) LRM + WS + TLI, (b) LRM + TLI + PP, (c) LRM + WS + PP, and (d) WS + TLI + PP, all applied on FSHS dataset. 75
Figure 4.8 Performance of the SVM model using the full features combination in terms of (a) ROC curve, and (b) The minimum classification error. 75
Figure 4.9 Performance of the NN model using the full features combination in terms of (a) ROC curve, and (b) Loss function. 76
Figure 4.10 ROC comparison between SVM and NN results using individual features models: (a) TLI, (b) PP, (c) WS, and (d) LRM applied on Arabic subset of ICDAR2013 dataset. 78
Figure 4.11 ROC comparison between SVM and NN results using individual feature models: (a) TLI, (b) PP, (c) WS, and (d) LRM applied on English subset of ICDAR2013 dataset. 79
Figure 4.12 AUC values of all features combination model applied on (a) Arabic subset and (b) English subset. 80

Figure 4.13 ROC curves and AUC values of (a) ResNet and (b) GoogleNet applied to the FSHS dataset 82
Figure 4.14 ROC curves and AUC values of (a) ResNet and (b) GoogleNet applied to the ICDAR2013 dataset 82
Figure 4.15 Comparison of feature importance between SVM and NN 85
Figure 4.16 ROC curves and AUC values using individual feature models: (a) SI, (b) PPI, (c) TLI, and (d) PWB, all applied to FSHS dataset 87
Figure 4.17 ROC curves and AUC values using three features combination models: (a) SI + PPI + TLI, (b) PPI + TLI + PWB, and (c) SI + TLI + PWB , all applied to FSHS dataset 88
Figure 4.18 Performance of the SVM model using all features combination applied to FSHS dataset in terms of (a) ROC curve, and (b) The minimum classification error 88
Figure 4.19 Performance of the NN model using all features combination applied to FSHS dataset in terms of (a) ROC curve, and (b) Loss function 89
Figure 4.20 ROC curves of SVM and NN results using individual features models: (a) SI, (b) PPI, (c) TLI, and (d) PWB applied on Khatt dataset. 91
Figure 4.21 ROC curves of SVM and NN results using all features models: (a) SVM and (b) NN, applied to Khatt dataset. 91
Figure 4.22 ROC curves and AUC results using individual features models: (a) TLI, (b) PPI, (c) SI, and (d) PWB applied on female and male subsets of the FSHS dataset. . 95
Figure 4.23 ROC curves and AUC values of the proposed age detection system of ALL features combinations models: (a) Male and (b) Female, applied on female and male subset of the FSHS dataset 96
Figure 4.24 ROC curves and AUC values of applying transfer learning to FSHS dataset using (a) ResNet and (b) GoogleNet 97
Figure 4.25 ROC curves and AUC values of applying Transfer learning using (a) ResNet and (b) GoogleNet to female and male subsets of the FSHS dtataset 98

Figure A.1 ROC curves and AUC values using two feature combination models: (a) PPI + PWB, (b) PPI + TLI, (c) SI + PPI, (d) SI + PWB, (e) SI + TLI, and (f) TLI + PWB, all applied to FSHS dataset 104
Figure B.1 ROC curves and AUC values of the proposed age detection system of two feature combination models: (a) TLI + PWB, (b) PPI + TLI, (c) SI + TLI (d) SI + PPI, (e) SI + PWB, and (f) PPI + PWB, applied on female and male subset of the FSHS dataset 107
Figure B.2 ROC curves and AUC values of the proposed age detection system of three feature combination models: (a) SI + TLI + PWB, (b) PPI + TLI + PWB, (c) SI + PPI + TLI, and (d) SI + PPI + PWB, applied on female and male subset of the FSHS dataset 108

List of Tables

Table 1.1 Comparison between Arabic, English and Hindi languages. 8

Table 2.1 Summary of different handwritten datasets reviewed found in the literature . 18

Table 2.2 A summary of gender detection systems found in the literature 23

Table 2.3 A summary of gender detection systems found in the literature 28

Table 3.1 Example of labelling five handwritten documents of the FSHS dataset 35

Table 3.2 Statistics of the number of the FSHS dataset writers in terms of gender, age, handedness and occupation 37

Table 3.3 Accuracy rates for different word spacing bin sizes 45

Table 3.4 Different accuracies for different margin lengths 48

Table 3.5 Different accuracy rates of different vertical and horizontal projection profiles 50

Table 3.6 Results of using different number of CCs on female handwriting 53

Table 3.7 Results of using different number of CCs on male handwriting 53

Table 3.8 Results of using different number of CCs on female and male handwriting . 54

Table 3.9 Different number of square results on female handwriting 55

Table 3.10 Different number of square results on male handwriting 56

Table 3.11 Different number of square results on female and male handwriting 56

Table 3.12 Different square dimension results on female handwriting 58

Table 3.13 Different square dimension results on male handwriting 58

Table 3.14 Different square dimension results on Female and male handwriting 58

Table 4.1 Gender detection confusion matrix 70

Table 4.2 Results of applying the proposed Gender Detection method to FSHS dataset . 71

Table 4.3 Results of applying the proposed gender detection system on Arabic subset of ICDAR2013 dataset. 77
Table 4.4 Results of applying the proposed gender detection system on English subset of ICDAR2013 dataset. 77
Table 4.5 Comparison between the results of our proposed gender detection method with other research methods applied on ICDAR2013 dataset. 80
Table 4.6 Results of applying transfer learning method to the FSHS dataset 81
Table 4.7 Results of applying transfer learning method to the ICDAR2013 dataset . . . 81
Table 4.8 Age detection confusion matrix 83
Table 4.9 Results of individual features models applied to FSHS dataset. 84
Table 4.10 Results of The best features combination models applied to FSHS dataset . . 84
Table 4.11 Results of the proposed age detection method applied to Khatt dataset 90
Table 4.12 Comparison between the proposed age detection method and the available works 92
Table 4.13 Results of the proposed age detection system applied to female and male subsets of FSHS dataset 94
Table 4.14 Results of applying the proposed automatic feature extraction method to the FSHS dataset 97
Table 4.15 Results of applying the transfer learning technique to the female and male subsets of FSHS dataset 97
Table A.1 Results of applying the proposed age detection system to FSHS dataset . . . 105
Table B.1 Results of the proposed age detection system applied to female and male subsets of FSHS dataset 109

Chapter 1

Introduction

1.1 Overview

Handwriting analysis (graphology) is the science of conducting personality traits, diseases, demographic properties and characteristics of an individual through his/her handwriting by measuring features such as slant, pen pressure, word spacing and many others. Handwriting is commonly known as brain writing because the writer can control what to write but not how to write. Therefore handwriting is a mirror reflecting the writer's characteristics.

Handwriting analysis is a powerful tool that can be used in any human activity where people interact, such as recruitment, management selection, physical and psychological health, security checking, career guidance, political ideology, personality profiling, social network analysis, medical diagnosis and forensic domain.

The interest in graphology goes back to 400 years. In 1622, Camilo Baldi wrote the first graphological essay which carried out systematic observations of handwriting, for what he was called the father of graphology. In 1897, the word graphology was derived from merging two Greek words graphein (to write) and logos (science) [1]. Another study was conducted in 1989, that aimed to learn how close a graphologist analysis of handwriting can come to the impression of the real person as perceived by people who know him well. The study summarized that graphological report

matches impressions held by other people better than chance [2]. Shamsuddin et al. [3] identified the relationship among five main selected graphology features such as proportion, calibre, baseline, slanting and space with the 16 personality factors in Cattll's 16PF. They found that the selected features and Cattall's 16 PF do have a correlation of 80.9%. In 2014, Miguel-Hurtado et al. [4]. studied the relationship between handwritten signature production and personality traits. Dynamic and static features were extracted from the signatures and compared with personality/biographic traits, where they found that there is a significant correlation between these aspects.

Research in handwriting analysis can be done by visually inspecting a set of relevant features. The manual way of handwriting analysis suffers from such drawbacks as the accuracy of the analysis depends mainly on the graphologist's skills, and it is costly and prone to fatigue. In addition, the analysis of graphologists might be influenced by the content of the handwriting. These disadvantages created the need to establish an automatic tool, which extracts the features and analyzes the writer behavior faster and more precisely with the help of a computer without human intervention [5].

The computational graphology process consists of several phases; image acquisition is the process of digitizing handwritten documents by scanning them using an optical scanner which consists of a transport mechanism and a sensing device which convert printed text, handwritten or an object to digital image. The acquisition can be either offline or online. The offline handwritten analysis utilizes an optical scanner or a digital camera to digitize a text document. However, in online handwriting analysis, an input tablet or light pen is used to capture the handwriting during the process of writing [6].

Preprocessing step is required to improve the quality of images before sending them to subsequent stages. Preprocessing is mainly divided into three steps, binarization, noise removal, and segmentation. Image binarization is the process of converting a given input gray-scale image into its binary representation. Global thresholding and local thresholding are some techniques that are used for binarization. Image noises are a common problem that occurs when scanning a document,

for many reasons, such as improper scanning process, imperfect paper, pen type and ink type that was used to write the document. So we need to remove these noises because they affect the performance of the handwritten document analysis system. The following are three types of noises that often appear in handwritten documents [7][8]. Background noise, which is a result of the imperfect paper or the inappropriate setting of the used scanner. The noise should be removed without losing the details of the handwriting. Thresholding techniques are usually used to remove the background noises. Rule lined noise, which appears because the handwritten is written on a ruled paper, so that the image may suffer from different problems, such as the lines may interfere with the text line. The algorithms cannot detect the broken lines. Many techniques can be used to detect ruled lines and remove them, such as Hough-transform and morphological operations. Salt and pepper noise, which appear in a document during the scanning process or due to dirt on the paper. They can be composed of one or more pixels, but usually, they are smaller than the size of the text objects. The median filter can be used to remove these noises.

Segmentation is an important step in preprocessing since it makes the processing of the handwritten image easier. The handwritten document can be separated into various segments, such as line segment, word segment, and letter segment.

In the feature extraction phase, a set of features such as slant, pen pressure, margins, baseline, word spacing and many others are extracted from the handwritten samples. The primary goal of feature extraction is to extract the most relevant information from a raw data and represent them in a reduced set of feature vectors, which makes it easier for the classifier to analyze the extracted features instead of the raw data, since using all candidates will easily lead to over-fitting. It is important to focus on the feature extraction phase because the performance of the classification system is highly dependent on the extracted features. The primary goals of feature extraction can be summarized as follows: extracting the most relevant information to increase the discrimination between the classes, extracting features to reduce the within class variability meanwhile, increase the between class variability, and discarding redundant information.

Finally, the classification can be done using many methods such as Neural Networks (NNs), Support Vector Machines (SVMs), rule-based approach and many others to develop the computational graphology for several languages. This stage aims to predict the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function (F) from input variables (X) to output variables (Y). Figure 1.1 shows the general structure of the computerized graphology system comprising the four main phases mentioned previously.

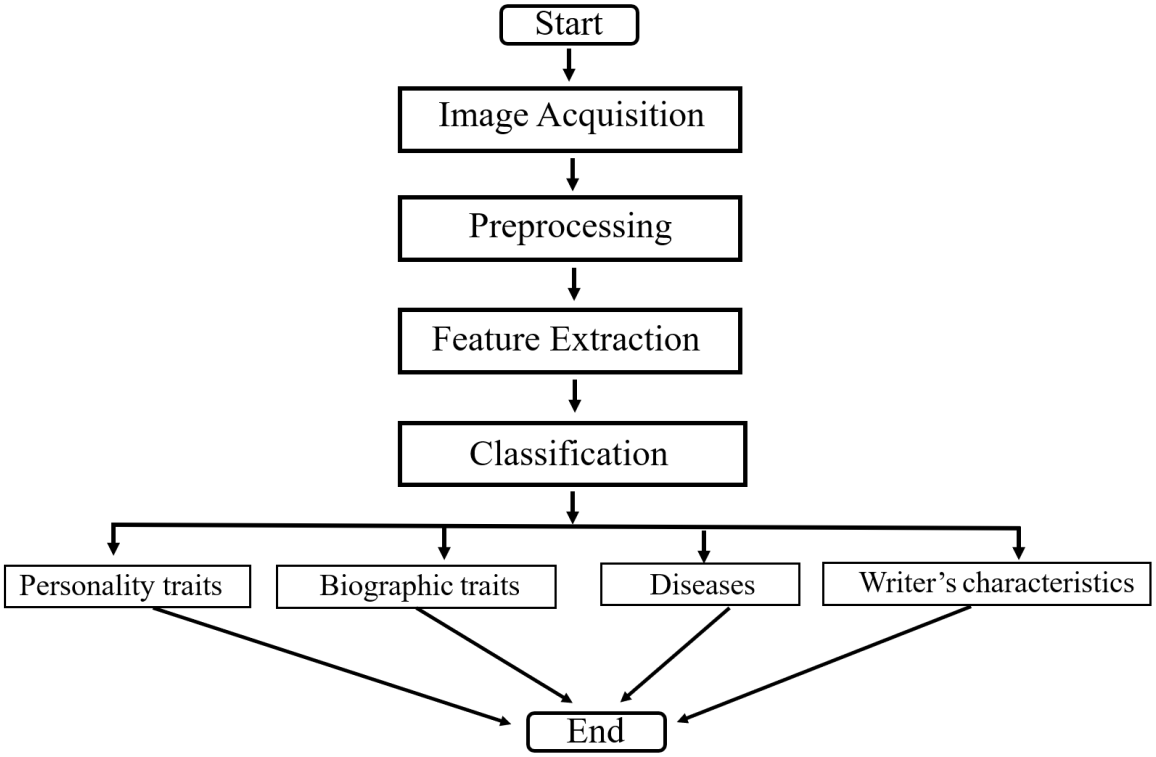


Figure 1.1: General structure of computerized handwriting analysis system

Our research is related to the discovery of biographic traits, particularly the gender and the age of the writer of a handwritten document automatically. Graphologists who analyze human handwriting have been able to detect gender from human handwriting. This is done by extracting some distinctive features. But, the manual detection process is expensive, tedious and exhausting, making the automation of the detection system an attractive alternative.

In our approach, we propose a new set of features to characterize the handwriting concerning gender and age of the writer. Characterizing powerful features makes it possible to distinguish between writers' categories. The classification of handwriting into gender (female and male) and age (youth and adult) has several applications. In forensic diagnosis, handwriting classification helps investigators focus on a specific group of writers. In medical diagnosis, it helps avoid developments in the detected disease. Two languages were used to conduct the proposed experiments, Arabic and English. We collected our Arabic handwriting dataset, which we named Free Style Handwriting Samples (FSHS). In the following section, we will explore the Arabic language characteristics, one of the biggest challenges in our work.

1.2 Arabic Language

The Arabic language is one of the most famous languages in the world. It is spoken by 186 million people in over 20 countries, especially in the Middle East and North Africa. It has many fonts such as: Alnasekh, Althulth, Alfarsi, Alruqa and Aldewani. Figure 1.2 shows different ways to write the same word with different Arabic font types. The analysis of Arabic language is far

Figure 1.2: Different fonts of Arabic language

more difficult to interpret compared to other languages such as Latin, Chinese, and French because of a range of challenges that characterized Arabic than other languages. Some of these challenges are: the dotting, direction of writing, diacritics, ligatures, the connectivity feature and the multiple

character shapes [9].

- **The dotting** [10], Some Arabic characters have exactly the same shape, but they differ by the addition of a group of dots or (zigzag) Hamza . To illustrate, dots in Arabic language are an integral part of a letter. They are very importatnt to distinguish between letters that have the same shapes but different sounds. Figure 1.3 shows how letters with the same shape can be differentiated only by the position of the dots.

Figure 1.3: Characters with the same shape but different dot positions

- **The direction of the writing** [11], Arabic language is written from right to left, whereas English language is written from left to right.

- **The diacritics** [9], The diacritics in the Arabic language is important in order to identify the meaning of the word. Some diacritics in Arabic language are named: fatha, kasra, skun, shadda and tanween. This can be seen in Figure 1.4.

Figure 1.4: Diacritics in Arabic language

- **The ligatures** [12], Ligatures and overlapping are some of the Arabic characteristics. Ligatures are atomic characters representing a compound of Arabic letters in some position of the word. Overlapping occurs when two or more characters overlap. Figure 1.5 shows some examples of ligatures in Arabic.

- **The connectivity feature** [12], Arabic text is written only in cursive (connected) mode. Characters in the same word are usually connected to each other, This can be shown in Figure 1.6.

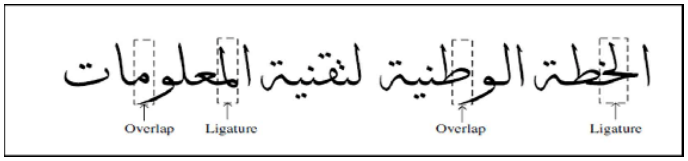


Figure 1.5: Ligatures in Arabic language

Figure 1.6: Connectivity in Arabic language

- **Multiple character shapes** [11], Arabic language consists of 28 primary characters, and each character can be written in different shapes depending on its position in the word (starting, middle, ending and isolated). Unlike Arabic language, English language has only two shapes for its characters capital and small letters. Figure 1.7 shows different shapes of the same letter based on its position of the word.

How Arabic Letters are Joined			
At the End of Words	In the Middle of Words	At the Start of Words	The letter
ـا | ا	ـا | ا	ا	ا
ـب | ب	ـبـ	بـ	ب
ـت | ت	ـتـ	تـ	ت
ـث | ث	ـثـ	ثـ	ث
ـج | ج	ـجـ	جـ	ج
ـح | ح	ـحـ	حـ	ح
ـخ | خ	ـخـ	خـ	خ
ـد | د	ـد	د	د
ـذ | ذ	ـذ	ذ	ذ
ـر | ر	ـر	ر	ر
ـز | ز	ـز	ز	ز
ـس | س	ـسـ	سـ	س
ـش | ش	ـشـ	شـ	ش
ـص | ص	ـصـ	صـ	ص
ـض | ض	ـضـ	ضـ	ض

Figure 1.7: Different letter shapes in Arabic language

Table 1.1 shows a comparison between the characteristics of Arabic, English and Hindi languages.

Table 1.1: Comparison between Arabic, English and Hindi languages.

	Characteristics	**Arabic**	**English**	**Hindi**
1.	Cursive	Yes	No	Yes
2.	Letter Shapes	4	2	1
3.	Diacritics	Yes	No	Yes
4.	Number of Letters	28	26	44
5.	Justification	R-L	L-R	L-R

1.3 Motivations

Automated applications identifying personal information without involving the subject is becoming a crucial research area. Therefore, automatic gender and age estimation or detection from handwritten documents is becoming an important research topic.

Gender detection is a bio-metric and forensic application. The gender can be detected from different bio-metric features such as physiological traits (face) or behavioral traits (handwriting) [13][14]. Despite the high similarity between individual writings, each person has his/her own writing preferences. Many researchers validated the relation between handwriting and gender detection what makes handwriting analysis for gender detection a vital research area in many sectors, including psychology, document analysis, paleography, graphology, and forensic analysis [15][16][17].

In the forensic domain, it helps to limit the investigation to a specific category of population. Additionally, processing each one of these categories separately can lead to improved results in writer identification and verification applications [16]. Psychology can also get benefits from research on handwriting style since it could be possible to identify correlations between the handwriting and some personality attributes of the writer [18]. In the field of Human-Computer Interaction, if the gender of a user can be automatically predicted, the computer applications could offer him/her a more personalized interaction (e.g., gender-oriented advertising). Bio-metric security can also benefit from handwriting prediction since this fact can be combined with other bio-metric modalities in

order to improve security when accessing computer systems [19].

Age detection from handwriting is essential in many applications, such as forensic studies and health diagnosis. In forensic studies, it helps limit investigation to a more targeted age group or category, which leads to improved results in writer identification. Similarly, in health diagnosis, age estimation or detection helps determine whether the development in handwriting is commensurate with age development. For example, detecting significant handwriting degradation in a child's handwriting indicates an early pathological and neuro degenerative state. In addition, distinguishing the normal evolution of handwriting with age from abnormal change is potentially related to cognitive decline [2][20].

Gender and age detection based on handwriting is a challenging and complex task for several reasons, for example the motor skills are not mastered before the age of 14. Moreover, the high similarity between people's writings made it harder to determine the right features to detect the writer's age and gender, either manually or automatically. These difficulties made the gender and age detection systems an active research topic. Research in these topics can be done by visually inspecting a set of relevant features such as slant, spacing and letter size. Meanwhile, it can be performed automatically by extracting features from offline or online handwriting. The manual way of detecting age and gender from handwriting suffers from many difficulties, such as being time-consuming, especially if there are many documents to be analysed. In addition, the process is expensive, tedious and exhausting [21]. Therefore, we focused on automating the gender and age detection systems from handwritten documents by proposing new powerful algorithms to automatically extract the most relevant features from his / her handwriting.

1.4 Objectives

Graphologists who analyze human handwriting have detected gender and age from human handwriting. This is done by extracting some features from the handwriting. The manual detection process is expensive, tedious and exhausting. This makes the automation of the detection system

important. In this research, we focused on automating the writer's gender and age detection systems from handwritten documents. To accomplish this purpose, we worked through a few steps. Figure 1.8 shows the main process followed to achieve the main objectives of this thesis.

Collecting and labelling a good dataset for training and testing any classification system is an essential step to achieve a good performance. Arabic is one of the oldest languages in the world. The Arabic characters are used to write Arabic, Farsi (Persian) and Urdu [22]. The Arabic language suffers from the absence of its software packages to help reduce this issue; we will train and test our proposed system using Arabic handwritten documents. More details are explained in section 3.1.

The performance of handwritten classification systems is highly dependent on the extracted features. It is fundamental to extract the most relevant information from the raw data and represent them in a reduced set of feature vectors [23]. To accomplish this, we proposed two main types of feature extraction methods. First, the Automatic feature extraction method using the transfer-learning concept. Second, the manual feature extraction method using image processing techniques. There is a specific type of feature suggested by the graphologists that can be extracted and analyzed to detect the gender and age of the writers. More details will be explained in section 3.3.

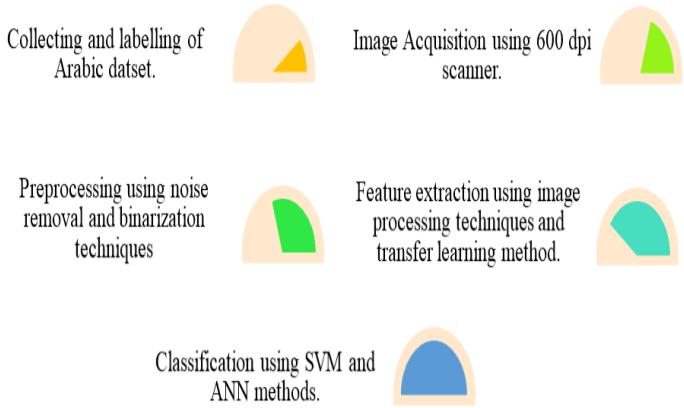


Figure 1.8: Main processes of the proposed system

1.5 Challenges

We faced many challenges during this research and they are summarized below:

- One of the biggest challenges in handwriting analysis systems is the lack of labelled databases to enable active innovation. Currently, as far as we know, there is no large and full text publicly labelled dataset for age detection applications either in the Arabic language or any other language.

- Another challenge in our work is dealing with the Arabic language because of a range of difficulties that characterize Arabic more than other languages, including the dotting, direction of writing, the diacritics, the ligatures, and the connectivity features. Image degradation, unexpected markings, and unseen writing styles also provide challenges.

- The few number of research in age detection from handwriting is another challenge faced our work since there is a limited background to build our system on.

- The high similarity in the handwriting between people is also a big challenge. So we put a considerable effort to set the most distinguishable gender and age related features to get a high classification rate.

We aim to provide a system that overcomes all these challenges and develop powerful gender and age detection systems from handwriting.

1.6 Contributions

This thesis focuses on developing and implementing gender and age detection systems from handwritten documents, Where the main contributions can be summarised below:

- We propose a new Arabic handwriting dataset, which can be used for building different handwriting classification and recognition systems.
- We propose a new set of novel features to detect both age and gender from handwritten documents by implementing new algorithms using different image processing techniques.
- New combinations of features are introduced. Several experiments were conducted to study the effect of each feature individually on the performance of the classification systems compared to the effects of the derived combinations.
- Transfer learning using two CNN architectures was applied to extract the features from the handwritten documents automatically.
- Experiments and evaluations were applied to compare the results of the two proposed feature extraction methods.

1.7 Thesis Outlines

This thesis organized as follows:

- Chapter 2 reviews and discusses previous work on gender and age detection from handwritten documents, as well as various datasets.
- Chapter 3 demonstrates the proposed framework in terms of data collection, preprocessing, feature extraction, and classification algorithms.
- Chapter 4 presents the proposed systems' outcomes. The employed datasets were described first, followed by the various evaluation methods. The results are then presented and discussed.
- Chapter 5 concludes the proposed works and suggests future directions.

Chapter 2

Literature Review

2.1 Datasets

Several databases have been used to research handwriting analysis and recognition. However, these databases differ in terms of the language used, number of writers, number of documents, and the way of collecting the samples, either online or offline. In the online type, the datasets are collected using a digitizer, where the x and y coordinates of the pen trajectory are recorded as a function of time while the writing is in process. IRONOFF [24] is an online database with 1000 English and French digitized texts written by 700 writers. It can be used for offline and online analysis systems. IAM dataset has an online version [25] which contains 1700 digitized documents written by 221 different writers. In addition, the CASIA [26] dataset has online Chinese and English digitized texts.

In the offline one, a scanned image that contains the text is the input. These datasets are more complex than online ones because only the completed text is available, which leads to poor knowledge of contextual information. What also makes it difficult is the lack of some prior information such as text position, size of text, order of strokes, stopping points and starting points. The presence of image noises is also one of the offline recognition system problems. The offline datasets can be categorised into two types: printed and handwritten.

There are many databases of the first type where the dataset can be obtained from different sources ranging from technical reports and business letters to newspapers and magazines. Some of the most used printed datasets are APTI [27], APTD/MF [28], and KAFD [29]. However, in our research, we are more concerned with handwriting, so in the following section, we will explore the most important offline handwritten databases.

CEDAR [30] (2002) dataset is one of the first large handwriting databases as it is a collection of handwritten English letters copied thrice by 1500 American writers. It was developed at the University of Buffalo to be used in text-dependent writer identification and verification. The database is not publically available.

IFN/ENIT [31] (2002) dataset contains 26459 handwritten Arabic Tunisian town/village names. Four hundred eleven writers were asked to fill a form with three columns and a text block at the bottom. The first column has 12 lines of printed Tunisian town/village names and postcodes. The other columns were filled with the names and the postcodes handwritten by the writer. 300 dpi scanner was used to scan all the documents. A pre-label, consisting of the postcode, the word as in Arabic code set ISO 8859-6, and a code for the Arabic character shapes, was assigned to each word. The dataset is useful for training and testing Arabic recognition systems and it is available for writer identification research purposes. However, the dataset contains only names of towns and villages of Tunisia, which led to a limited variety in the Arabic text.

IAM [32] (2002) dataset consists of handwritten English sentences derived from the electronic version of the Lancaster-Oslo/Bergen (LOB) corpus. The dataset contains 1066 forms with printed text taken from the corpus. Four hundred writers were asked to rewrite the printed text by hand on the forms. The total number of words is 82227, distributed over 9285 lines. The documents were scanned using a 600 dpi scanner. The dataset is available for researchers upon request. It can be used in different research areas such as handwriting recognition systems based on linguistic knowledge beyond the lexicon level, segmentation problems, writer identification, and machine-printed

text processing. The dataset is publicly available.

AHDB [33] (2004) dataset is an Arabic collection of words used to represent numbers on checks and some Arabic sentences used in writing checks. It also has some handwritten pages of randomly chosen text by each writer. The documents are divided into 105 folders, with 315 documents written by 105 volunteers. The number of writer of this dataset is considered small which limits diversity.

RIMES [34] (2006) dataset is a French database consisting of handwritten mails written by 1300 persons. The total number of dataset documents is 12723 pages corresponding to 5605 letters of 2-3 pages. Several competitions have been held using this dataset, such as ICFHR2008, ICDAR2009, and ICDAR2011. The dataset is not publicly available.

MADCAT (2009) [35] dataset is a collection of 1000 Arabic handwritten documents of 325 writers. The documents were scanned using 600 dpi resolution. Unfortunately, it is not publicly available.

CASIA [26] (2011) consists of Chinese handwritten characters and continuous scripts datasets. The name CASIA is related to the builder of the dataset, the Institute of Automation of the Chinese Academy of Sciences. The dataset is divided into two categories: offline and online. 1020 writers were asked to utilize Anoto pen on papers such that both datasets were written simultaneously, each of the datasets contains six datasets. Three of them are for isolated characters, and three are for handwritten texts. The isolated character dataset contains 3.9 million samples with 7356 classes, while the continuous text dataset is built of 5090 pages and 1.35 million isolated characters. Both datasets are helpful for many Chinese handwritten document analysis tasks.

Qatar University Writer Identification (QUWI) [36] (2012) is an offline Arabic and English Handwriting dataset. It contains 4068 documents written by 1017 volunteers. The writers were asked to copy a particular text and generate a random text. Moreover, they were asked to write their name, age, gender, handiness, professions, educational level and nationality. The dataset consists of

60000 words written in Arabic and English for text-independent analysis. At the same time, it has more than 100000 words for both languages written by the same 1017 writers for text-dependent analysis. The data set documents were scanned using 600 dpi resolution. The ages of the writers range from 12 years old to older than 40 years. Females wrote 52% of the documents, and males wrote the rest. It can be used in many writer identification applications such as gender, age and handiness detection.

CVL [37] (2013) dataset is a German and English collection of text handwritten by 311 writers. It contains 2160 handwritten documents scanned using 300 dpi resolution. CVL dataset was used in ICDAR2011 and ICFHR2012 contests.

Kaggle (ICDAR 2013) [38] dataset consists of 1900 images written by 475 writers in Arabic and English. Each writer wrote two English documents and two Arabic documents. The first page of both languages has a text which varies from one writer to another, but the second page contains the exact text for all writers. The dataset documents were scanned using 600 dpi resolution. Kaggle dataset was used for the competition on gender prediction from handwriting in the 12th International Conference on Document Analysis and Recognition (ICDAR 2013).

LAMIS-MSHD [39] (2014) is a multi-script offline handwriting database. One hundred randomly selected Algerian volunteers were asked to write 600 Arabic samples, 600 French text samples, 1300 signatures and 21000 digits. In addition, each writer wrote 13 forms, making 1300 forms. The documents were scanned using 300 dpi. The dataset can be used in multi-script writer identification and verification applications.

Khatt [40] (2014) dataset consists of 2000 randomly selected Arabic paragraphs from 46 sources. All paragraphs have a total of 9327 text lines handwritten by 1000 writers. The documents were scanned using 200, 300, and 600 dpi resolutions. It was randomly divided into 70%, 15%, and 15% sets for training, testing, and validation, respectively.

ALTID [41] (2015) dataset is an Arabic/Latin handwritten text database. Seven hundred thirty-one printed Arabic and Latin pages with gray-scale format were scanned using 300 dpi resolutions. The dataset consists of 460 Arabic and 582 Latin text blocks, which 17 volunteers wrote.

AWIC2011 [42] dataset written by 54 writers, where each writer wrote three paragraphs. It has a relatively small number of writers. The CV-MUSCIMA [43] dataset contains 1000 music sheets written by 50 different musicians. It can be used for the identification of musicians. The BBN [44] dataset includes 39500 documents written by 259 volunteers, which is not high enough. AL ISRA [45] database was collected randomly from 500 students of AL-Isra University in Amman, Jordan. It contains 37000 Arabic words, 10000 digits, 2500 signatures, and 500 free form Arabic sentences. Therefore, it has a limitation regarding Arabic text because it does not have entire Arabic paragraphs of text. IIT-Demokritos [46] is a Latin dataset consisting of 208 documents written by 26 writers.

Many factors determine which of these datasets is most suited to the study's objectives. The number of writers, the quantity of documents, the language used, and the database category (words, characters, paragraphs, or digits) are the most essential aspects. Furthermore, the dataset's availability, and finally, the presence of appropriate ground truth connected to the research topic. Table 2.1 shows a summary of the databases covered previously.

Our work is related to gender and age detection from handwritten documents. The QUWI dataset was the most convenient and effective one to employ in our research. We tried contacting the authors to to get the the ground truth, but we didn't hear back, so we made the decision to make our own Arabic dataset instead. So we created our Arabic handwritten dataset to train and evaluate both proposed approaches. It was collected in Amman, the capital city of Jordan. We called it Free-Style Handwritten Samples (FSHS) [47]. More details about the FSHS dataset is available in section 3.1. Two different datasets were chosen to compare the performance of the proposed approaches with other research results. First, QUWI-ICDAR2013 dataset was used to conduct the gender detection experiments. Second, the Khatt dataset was used for age detection experiments. More details can be found in section 4.0.1.

	Dataset Name	Language	Number of Writers	Number of Documents
1	CEDAR	English	1500	1500
2	IFN/ENIT	Arabic	411	2200
3	IAM	English	657	1539
4	AHDB	Arabic	105	315
5	RIMES	French	1300	12723
6	MADCAT	Arabic	325	1000
7	CASIA	Chinese	1020	5090
8	QUWI	Arabic, English	1017	5085
9	CVL	German, English	311	2160
10	ICDAR2013	Arabic, English	475	1900
11	LAMIS-MSHD	Arabic, French	100	1300
12	Khatt	Arabic	1000	2000
13	ALTID	Arabic, Latin	17	460 Arabic, 582 Latin
14	AWIC2011	Arabic	54	161
15	CV-MUSCIMA	Music scores	50	1000
16	IIT-Demokritos	Latin	26	208
17	BBN	Arabic	259	39500
18	IRONOFF	English, French	700	>1000
19	**Our dataset (FSHS)**	**Arabic**	**2200**	**2700**

Table 2.1: Summary of different handwritten datasets reviewed found in the literature

2.2 Gender Detection

Interest in gender detection began many years ago, when in 1999, Huber and Hadrick [48] presented the correlation between handwriting and several categories, including age, handedness, gender, and ethnicity. However, unfortunately, the old-fashioned way of gender classification suffers from some drawbacks, such as the accuracy of the analysis depends on the analyst's skills, its costs and prone to fatigue. These disadvantages created the need to establish an automatic tool, which analyzes the handwriting and detects the writer's gender faster and more precisely with the help of a computer without human intervention [49]. In the following section, we are going to explore some recent automatic handwriting gender classification systems. Table 2.2 summarises the gender detection systems mentioned in the literature.

Numerous feature types have been extracted from handwritten documents to classify the gender

of the writer successfully. Some of the classification systems depend on the extraction of geometric features from the handwriting. Almadeed and Hussain [50] proposed an offline automatic handwriting prediction system of age, gender, and nationality based on a set of geometric features extracted from the handwriting images. They combined these features using random forests and kernel discriminative analysis. Their system classification rates were 74.05%, 55.76%, and 53.66% for gender, age, and nationality classification, respectively. Bi et al. [51] presented a new approach for gender detection from handwriting named kernel mutual information (KMI) which focuses on using multi-features to identify the gender of the writer. They extracted geometrical features including slant orientation, curvatures, and transformed features such as Fourier features and WD-LBP features. They used SVM to classify the gender of the writer. Their system achieved an accuracy of 66.3% on ICDAR2013 dataset and an accuracy of 66.7% on the registration-document-form (RDF) dataset.

Seddiqi et al. [52] proposed an offline prediction system to classify the gender from handwriting images by extracting a set of features such as slant, curvature, texture, and legibility. Artificial neural network (ANN) and support vector machine (SVM) were used for the classification task. They applied their method on two different datasets: QUWI [36] and MSHD [39]. Their approach achieved classification rates of 68.75% and 67.50% when applying SVM and ANN on QUWI, respectively. Moreover, they produced a classification accuracy of 73.02% and 69.44% when applying SVM and ANN on MSHD dataset, respectively. Maken and Gupta [53] extracted a set of features related to the physical appearance of the handwriting, such as slants, area, and perimeter. The authors achieved a classification rate of 61.5%, 63.3% and 56.5% when applying their system on the Kaggle (ICDAR2013) dataset using SVM, logistic regression, and KNN, respectively. Dargan and Kumar [54] proposed an offline gender detection system to extract a set of features and use several classifiers. The approach was evaluated based on two factors, the accuracy and the CPU elapsed time after and before applying PCA. Zoning, diagonal, transition and peak extend are the main extraction feature methods of the system, achieved accuracy rates of 85.85%, 89.85% and 91.% using a decision tree, random forest and gradient boosting classifiers, respectively.

Textural features also caught the attention of researchers. Akbari et al. [55] proposed a gender detection system which relies on a global approach that considered the writing images as textures. A probabilistic finite-state automaton was used to generate the feature vectors from a series of wavelet sub-bands. They used SVM and ANN for the classification phase, and they used QUWI and MSHD to evaluate their work. Up to 80% accuracy of classification was achieved in this work. Gattal et al. [49] proposed a gender detection system that uses a combination of different configurations of basic oriented image features to extract the texture features from the handwritten samples. They used an SVM to analyze the extracted features and reveal the gender of the writer. Their proposed method achieved classification rates of 71.37%, 76%, and 68% for the ICDAR2013, ICDAR2015, and ICFHR2016 experimental settings, respectively.

Gornate et al. [56] extracted textural features from handwritten signature images. The proposed system was evaluated on the authors own dataset. They got accuracy rates of 96.17%, 98.72%, and 100% for K-NN, decision tree and SVM, respectively. Mirza et al. [14] used Gabor-filter to extract the textural features from the handwriting images, to distinguish between male and female writings. The classification task in this work was carried using a feed-forward neural network. They also used QUWI dataset to train and evaluate their method, where they divided the experiments into four tasks. In the first one, they used only Arabic samples for training and testing, while in the second task, they used only English writings for training and testing. However, in the third task, they used Arabic samples for training and English samples for testing. The fourth task was the opposite of the third task, where they used English samples for training and Arabic samples for testing. Their method achieved accuracy rates equal to 70% for task1, 67% for task2, 69% for task 3, and 63% for task 4, respectively.

Fahimeh and Alirezal [57] proposed a spatial pyramid matching technique to detect the gender of the writer as the input document is divided into sub-regions. A histogram of each region is then calculated. Finally, SVM and ensemble classifiers were used to determine whether the writer is female or male. QUWI and MSHD benchmark datasets were used to evaluate the performance of the proposed approach. As a result, the system got accuracy rates of 69% and 62% using SVM and

ensemble classifiers, respectively. Abbas et al. [58] presented a comparative study to evaluate different local-binary methodologies in the classification of gender from handwritings. They proposed a system that presents the images using low-pass filtering features, and then they analyzed these features using the SVM technique. The authors evaluated their method using the QUWI dataset. They achieved 72% of classification accuracy.

Recently, deep learning becomes a powerful tool for image classification and automatic feature extraction. Illouz et al. [59] proposed a novel deep learning approach for an offline handwritten gender classification system. They used a convolution neural network (CNN) which extracted the features automatically from the handwriting images and classified the gender of the writer. The authors created their own dataset, which consists of Hebrew and English handwriting samples of 405 participants. They evaluated their system in two categories: first, Hebrew-English classification, where they trained the system on Hebrew samples and tested it on English samples, and they got a 75.65% of classification accuracy. While in the second category, they trained the system on English samples and tested it on Hebrew samples. The accuracy rate for this category was 58.29%. Morera et al. [60] also used a convolution neural network to detect gender and handedness from offline handwriting. They created their own CNN architecture to classify the handwriting into four multi-classes, including left-handed woman, right-handed woman, left-handed male, and right-handed male. The system achieved an accuracy of 80.72% for gender classification on IAM dataset and an accuracy of 68.90% on Khatt [40] dataset. Xue et al. [61] used attention-based two-pathway Densely Connected Convolutional Networks (ATP-DenseNet) to detect the gender of a given handwritten document writers. They extracted the handwriting features in two pathways. First, the ATP-DenseNet, which extracts hierarchical page features. Second, attention-based DenseNet (A-DenseNet) extracts the word features. Their work produced classification rates of 65.2%, 77.6% and 74.1% when applied on ICDAR2013, IAM and Khatt datasets, respectively.

Rabaev et al. [62] used the transfer learning technique to extract fixed features from handwritten documents. The system was applied to the ICDAR2013 dataset and got an accuracy rate of 67% when using ResNet50. Youssef et al. [63] proposed an offline gender detection system

based on language-independent analysis. The authors applied their method on both subsets of the ICDAR2013 dataset (Arabic and English). They used an SVM classifier for the proposed analysis, and they achieved accuracy rates of 68.6% and 85.7% on the aforementioned subsets, respectively.

The signature of the writers received the attention of some researchers to classify gender. Maji et al. [18] used handwriting signatures to detect the gender. They extracted several features from the signature images, including roundness, skewness, kurtosis, mean, standard deviation area, Euler number, the distribution density of black pixels, and the connected components. They used a back-propagation neural network (BPNN) to analyze these features. Their system achieved an 80.7% accuracy with Euler number feature and a 77% accuracy without the Euler number feature. Djeddi et al. [64] proposed an offline prediction system to classify the writer's gender from handwriting images by extracting a set of features such as slant, curvature, texture, and legibility. Artificial neural network (ANN) and support vector machine (SVM) were used for the classification task. They applied their method on two different datasets, QUWI [36] and MSHD [15]. Their approach achieved classification rates of 68.75 and 67.50 when applying SVM and ANN on QUWI, respectively. Moreover, they produced a classification accuracy of 73.02% and 69.44% when applying SVM and ANN on MSHD dataset respectively.

Online features have been extracted using special devices such as tablets. Nogueras et al. [13] presented a gender classification schema based on online handwriting. The authors used a digital tablet to extract online features related to the pen trajectory, such as pen-up, pen-down, pen pressure. They achieved a classification accuracy rate of 72.6%. Mukherjee et al. [65] proposed an online Bangla vowel modifiers gender detection system. A CNN was applied to the handwritten documents of 27 writers. The system achieved a classification rate of 91.15%.

Navya et al. [66] proposed an offline gender detection system that relies on extracting multi-gradient features from the scanned handwritten documents. They evaluated their method using IAM [32] and QUWI datasets. They achieved 75.6% of classification accuracy. Maken et al. [19] provided a comprehensive review of various techniques used for gender detection systems. Gattal et

al. [67] proposed a new method for gender detection from handwriting based on extracting cloud of line distribution (COLD) and Hinge features. They used 1000 handwritten samples from the QUWI dataset to evaluate their system. They used 500 samples for training, 250 samples for validation, and 250 samples as a test set. They got classification rates of 71% and 68% for both Hinge and COLD features, respectively.

Several competitions have been held to compare the performance of gender detection systems on different languages, such as the International Conference on Document Analysis and Recognition (ICDAR2013 [38], ICDAR2015 [64]) and International Conference on Frontiers in Handwriting Recognition (ICFHR2012 [68], ICFHR2016 [69]).

Table 2.2: A summary of gender detection systems found in the literature

	Authors	Features extracted	Database used	Classification rate
1.	Almadeed and Hussain [50]	Geometric features	QUWI dataset (70% training, and 30% testing).	74.06% of gender classification accuracy
2.	Djeddi et al. [64]	Slant, curvature, legibility	QUWI dataset (1200 samples were used for training, and 700 samples were used for testing). while they used documents of 42 writers from MSHD dataset for training and the same number for testing.	73.02% and 69.44% of accuracy when they applied SVM and ANN on QUWI and MSHD datasets, resepectively.
3.	Maji et al. [18]	Signature features: roundness, skewness, mean, standard deviation, Euler number, black pixels distribution density	They used their own database(500 samples)	80.7% of accuracy with Euler feature and 77% of accuracy without Euler feature.
4.	Nogueras et al. [13]	Online features related to pen-trajectory	They used their own online dataset where 100 samples were used to build the models, and 242 samples were used for testing.	72.6% of accuracy
5.	Mirza et al. [14]	Textural features	QUWI dataset was distributed as 300 samples for training, and 200 samples for testing.	They got 70% of classification accuracy, when they used only Arabic samples for training and testing

Continued on next page

Table 2.2 – continued from previous page

	Authors	Features extracted	Database used	Classification rate
6.	Akbari et al. [55]	Textural features	They used 1900 samples from QUWI dataset for training and testing, and a total number of 1008 samples of MSHD dataset for training and testing.	Up to 80% of classification rate.
7.	Navya et al. [66]	Multi-gradient features	They used 990 handwritten samples from IAM and QUWI datasets	75.6% of classification accuracy
8.	Gattal et al. [49]	Textural features	ICDAR2013 (564 samples were used for training and 386 samples were used for testing), ICDAR2015 (300 samples were used for training and 200 samples were used for testing) and ICFHR2016 (500 samples were used for training and 500 samples were used for testing)	They got classification rates of 71.37%, 76% and 68% when they applied their method on the aforementioned datasets, respectively.
9.	Illouz et al. [59]	They used deep learning	They used their own dataset (810 handwritten samples)	They got 75.65% of classification accuracy, when they trained and tested their method on Hebrew and English samples, respectively.
10.	Morera et al. [60]	They used deep learning	IAM and Khatt datasets (8462 samples were used for training and 2464 samples were used for testing).	They got an accuracy rate of 80.72% and 68.90%, when they applied their method on IAM and Khatt datasets, respectively.
11.	Maken et al. [19]	Geometrical features including, slant orientation, curvatures, and transformed features including, Fourier features and WD-LBP features.	registration-document-form (RDF) and ICDAR2013 dataset (500 samples were used for training and 300 samples were used for testing)	Their system achieved an accuracy of 66.3% on ICDAR2013 dataset and accuracy of 66.7% on the registration-document-form (RDF) dataset.
12.	Gattal et al. [67]	Cloud of line distribution (COLD) and Hinge features	QUWI dataset. They used 500 samples for training, 250 for validation, and 250 for testing.	Their system achieved accuracy rates of 71% and 68% for both Hinge and COLD features, respectively,
13.	Bi et al. [51]	Slant, curvature and chain code	Arabic, English and Chinese	66.7%
14.	Abbass et al. [58]	Textural features	QUWI dataset	69% for script-independent and 71% script-dependent

Continued on next page

Table 2.2 – continued from previous page

	Authors	Features extracted	Database used	Classification rate
15.	Siddiqi et al. [52]	orientation, curvature, text based and fractional features	QUWI and MSHD	68.75% with QUWI dataset and 73.02% with MSHD dataset
16.	Maken and Gupta [53]	Slant, area and perimeter	564 samples of 282 writers with 2 samples per writer	60.50% and 56.5% for SVM and K-NN, respectively.
17.	Dargan and kumar [54]	Zoning, transition, diagonal and peak extend based are used in the feature extraction methods	Gurumukhi characters (35000 samples of male handwriting and 35000 of female handwriting)	85.85% for decision tree, 89.85% for random forest and 91.4% for extreme gradient boosting. All these results before applying PCA.
18.	Gornale et al. [56]	Textural and statistical features (LBP and HOG)	Their own signatures dataset of 4790 samples	96.17%, 98.72% and 100% for K-NN, decision tree and SVM, respectively
19.	Fahimeh and Alirezal [57]	A set of histograms of each sub region of the image are concatenated to form the feature vector	MSHD (600 Arabic and 600 French samples) and QUWI (950 Arabic and 950 English samples)	They got classification rates of 73% and 70% using SVM on the QUWI dataset. In addition to 69% and 68% using SVM on the MSHD dataset.
20.	Xue et al. [61]	Hierarchical page features and word features	ICDAR2013 (1128 samples were used for training and 772 were used for testing), IAM (2573 samples were used for training and 1286 were used for testing) and Khatt (1333 samples were used for training and 667 were used for testing)	classification rates of 65.2%, 77.6% and 74.1% when applied on ICDAR2013, IAM and Khatt datasets, respectively.
21.	Rabaev et al. [62]	Transfer learning was used to extract automatic features	ICDAR2013 (475 writers (221 males, and 254 females)) and Hebrew dataset (819 samples) 80% of the dataset was used for training, 10% for testing and 10% for validation.	67% and 50% accuracy rates were obtained after applying ResNet50 on the ICDAR2013 English and Arabic subsets, Respectively.
22.	Mukherjee et al. [65]	Online features related to pen-trajectory	Bangla vowel modifiers handwritten by 27 writers	91.15% using CNN
22.	Youssef et al. [63]	Textural and directional features	ICDARR 2013 (Arabic and English)	68.6% for Arabic and 85.7% for English

2.3 Age Detection

Few researchers in the literature have studied the problem of automatic age detection from handwriting. These studies vary in age groups to be detected and the extracted features proposed to detect age. Table 2.3 summarises the age detection systems mentioned in the following literature.

Basavaraja et al. [70] proposed a new method to estimate age from handwriting based on extracting disconnectedness features using Hu invariant. First, they explored the intersection points in the cany edge images. Then, they used K-means clustering to classify the handwritten documents of their dataset into four classes. Class -1 with ages ranging from $11 - 12$ years, class -2 with ages ranging from $13 - 16$ years, class -3 with ages ranging from $17 - 20$ years and class -4 with ages ranging from $21 - 24$ years. Each class has 100 images. However, for the two public datasets, IAM [32] and Khatt [40], the images are divided into two classes. As a result, an accuracy rates of 66.25% has been achieved using the IAM dataset and 64.44% using the Khatt dataset.

Marzinotta et al. [20] provided a method to classify age and gender from online handwriting features available in the IRONOFF [24] dataset in French and English languages. Their work is a two-layer schema. They used global features of handwritten words for a writer independent clustering in the first layer. However, the second layer was used to determine the stability across words for each writer by clustering the style variation at the previous level.

Bouadjenek et al. [21] introduced two gradient features to classify the writer's age, gender and handedness: the histogram of oriented gradients and gradient local binary patterns. They used the Support Vector Machine (SVM) [71] method to classify the documents. IAM and Khatt datasets were used to evaluate the system. IAM dataset was divided into two age groups based on age ranges: $25 - 34$ years and $35 - 56$ years. On the other hand, the khatt dataset was divided into three ranges that can be defined as "under 15 years old", "$16 - 25$ years old", and "$26 - 50$ years old". They got a 70% accuracy rate when using the IAM dataset and a 55% accuracy rate when using Khatt dataset.

Almaadeed and Hassaine [72] developed a handwriting analysis system for age, gender, and nationality classification. They extracted a set of geometrical features combined using random forests and kernel discriminant analysis. They applied their system OUWI [36] dataset, and they got accuracy rates of 55.7% and 60.62% for age detection when all writers produced the same handwritten text and when each writer produced different handwritten texts, respectively.

Marzinottto et al. [2] proposed an online age classification system based on a two-level clustering scheme. At the first level, writer independent word clusters are generated. On the other hand, the second level generates a bag of prototype words from each writer's words. Supervised learning then is used to categorise the handwritten documents in terms of age. A dataset sample acquired from Broca Hospital in Paris was used to conduct the experiments. The writer ages range between 60 and 85 years old. Their approach came out with the following findings: first, people above 65 years old present three handwritten patterns regarding the dynamic features, pen pressure and time on air. Second, people aged above 80 years have almost the same unique style with lower speed.

Table 2.3: A summary of gender detection systems found in the literature

	Authors	Extracted features	Database used	Age range	Classification
1	Basavaraja et al. [70]	Disconnectedness features	IAM and Khatt datasets	Two ranges for IAM dataset : 25-34 years old. 35-65 years old. two ranges for the Khatt dataset: 16 to 25 years old 26 to 56 years old	66.25% and 64.44% for IAM and Khatt datasets, respectively.
2	Marzinotta et al. [20]	Online features related to pen trajectory	IRONOFF (French and English)	Six age ranges: 7-13 years old 14-19 years old 20-30 years old 31-45 years old 46-59 years old 60-77 years old	satisfactory results

Table 2.3 – continued from previous page

	Authors	Extracted features	Database used	Age range	Classification
3	Bouadjenek et al. [21]	Gradient features	IAM and Khatt datasets	Two ranges for IAM datase : under 35 years old more than 35 years old two ranges for Khatt dataset: 16 to 25 years old 26 to 50 years old	70% and 55% for IAM and Khatt dataset, respectively.
4	Almaadeed and Hassaine [72]	Geometrical features	QUWI	Seven age ranges: 1950 to 1965 years 1966 to 1975 years 1976 to 1985 years 1986 to 1990 years 1991 to 1995 years 1996 to 2000 years 2001 to 2012 years	55.7% for text dependent, and 66.62% for text- independent

Table 2.3 – continued from previous page

	Authors	Extracted features	Database used	Age range	Classification
5	Marzinottto et al. [2]	Online features related to pen trajectory	IRONOFF (French and English)	Six age ranges: 11-17 years old 18-35 years old 36-50 years old 51-65 years old 66-75 years old 76-86 years old	satisfactory results

Chapter 3

Proposed Framework

In this research, we present a novel database of Arabic handwritten documents. Each sample is labelled to show the age and the gender of the writer. Also, we propose machine learning-based detection systems by extracting a set of features recommended by a psychologist and a graphologist using image processing techniques. In addition, a transfer learning method is used to extract the features automatically. A comparison between the two types of features is also conducted.

The automatic handwriting analysis process consists of several phases; Image Acquisition, Preprocessing, Feature extraction and Feature analysis. This section describes the proposed methodologies concerning the four mentioned stages. Figure 3.1 shows the diagram of the proposed detection systems.

3.1 Dataset Preparation and Image Acquisition

Several databases have been used to conduct research in handwriting analysis and recognition. However, these databases differ in terms of the language used, the number of writers and the number of documents. We created our Arabic handwritten dataset to train and evaluate the proposed approach. It was collected in Amman, the capital city of Jordan. We called it the Free-Style Handwritten Samples (FSHS) [47].

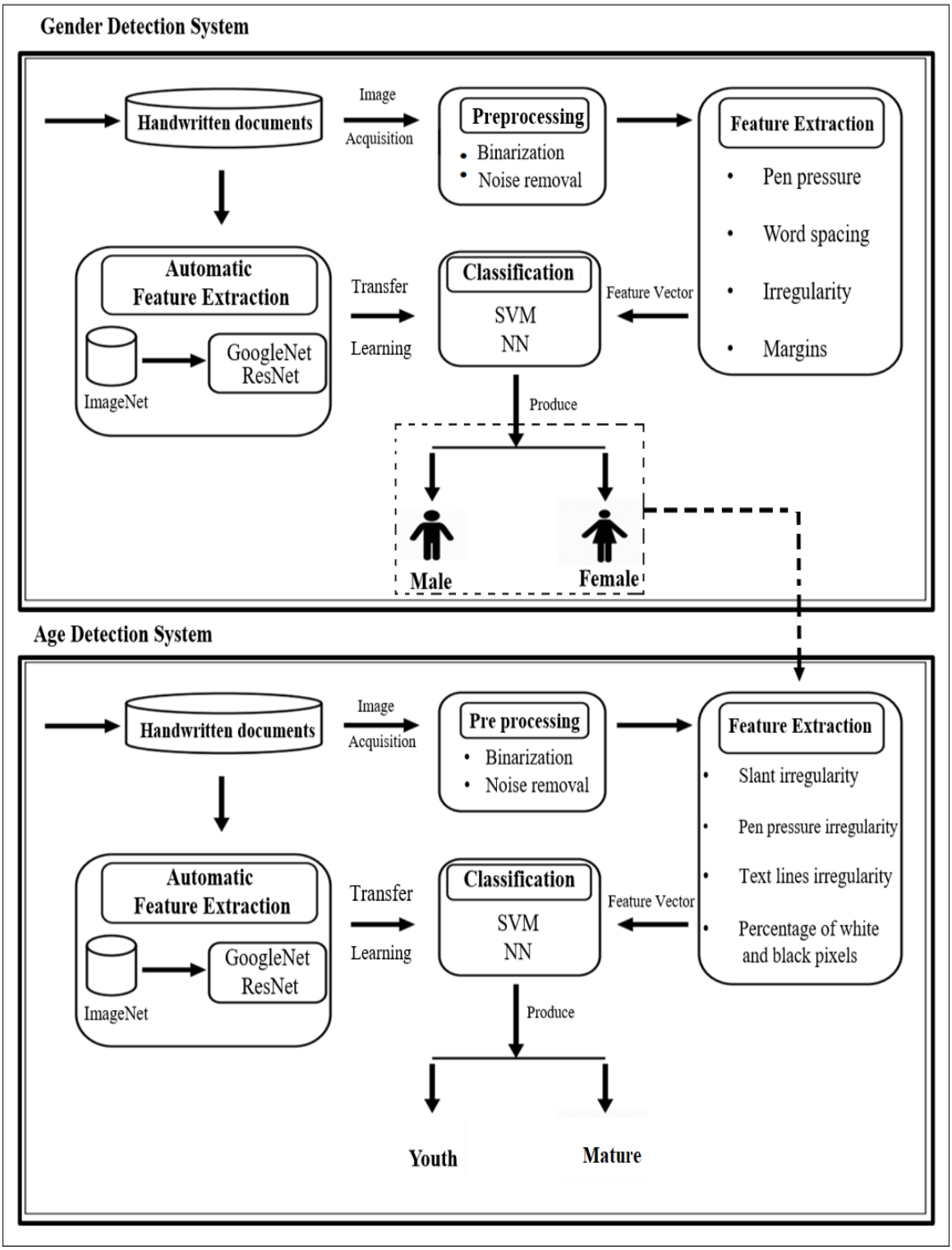
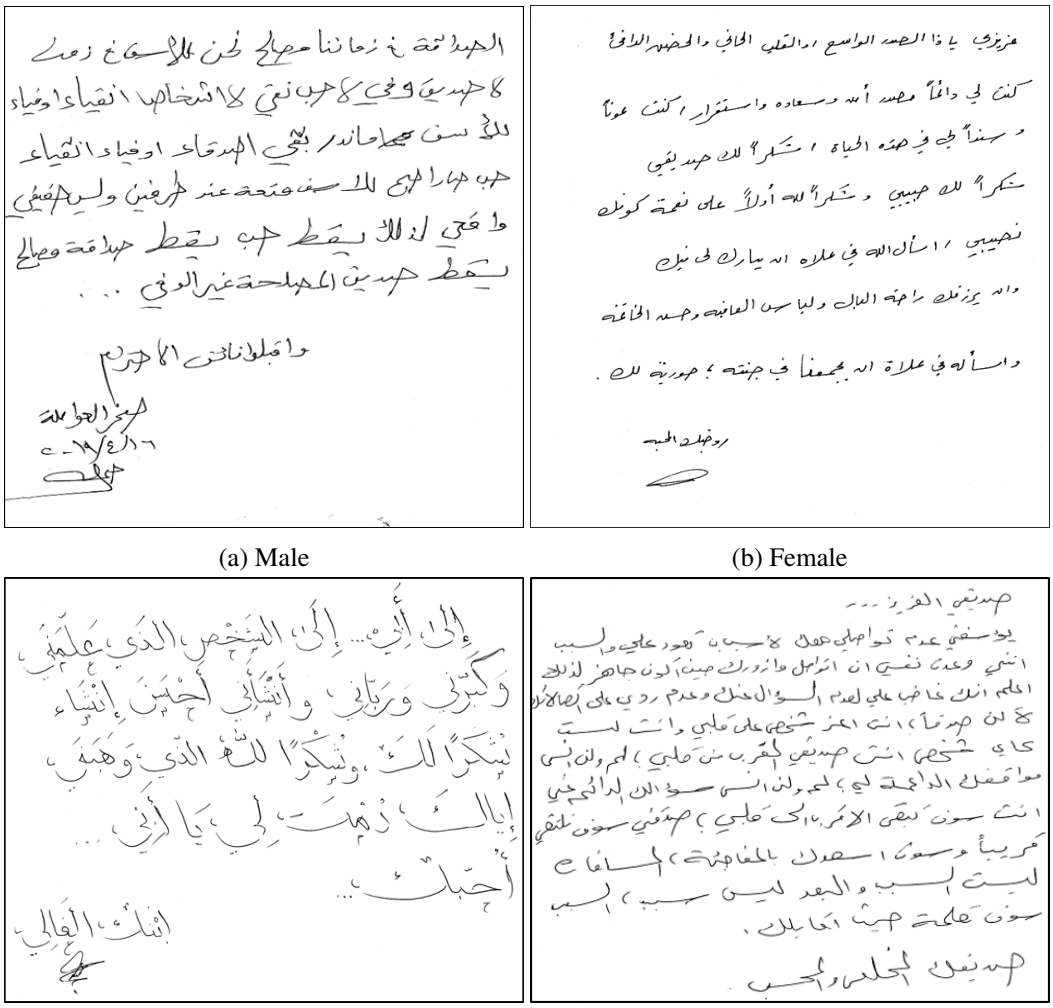


Figure 3.1: General structure of the proposed systems

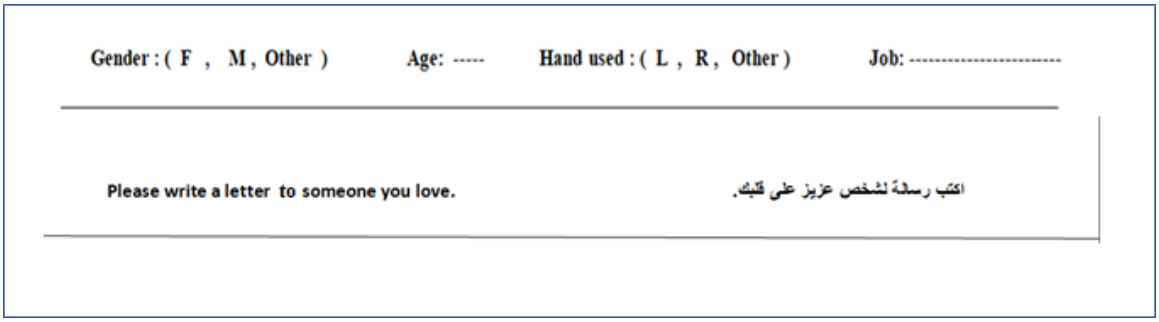
Two thousand two hundred volunteers were asked to write a letter to someone they love, to be large enough to cover all the different handwritten styles. In addition, some volunteers were asked to copy some paragraphs. No conditions were put on the type of instrument used for writing to ensure the writers' convenience. However, we provided them with a white sheet of paper to write their letter on it. The dataset consists primarily of text-independent documents but about 500 of them text-dependent samples. For example, Figure 3.2 shows some text-independent specimens concerning gender and age, while in Figure 3.4, the two samples are text-dependent.



(a) Male (b) Female

(c) Youth (d) Adult

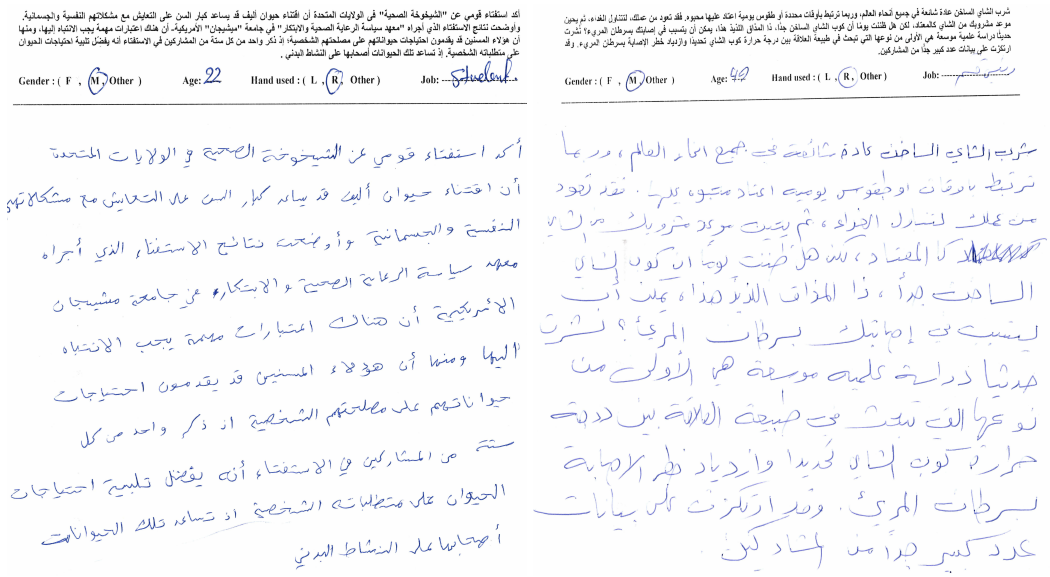
Figure 3.2: Samples of the dataset (FSHS) in terms of gender and age



Gender : (F , M , Other) Age: ----- Hand used : (L , R , Other) Job: ----------

Please write a letter to someone you love. اكتب رسالة لشخص عزيز على قلبك.

Figure 3.3: Information page of samples in the dataset



(a) Text1 (b) Text2

Figure 3.4: Text dependent samples of FSHS dataset

The writer's ages range from 15 to 75 years old. The variation in the educational levels of the participants is also interesting. Indeed, the volunteers included: secondary school students and university students as well as workers, employees, engineers, doctors, professors, accountants, and secretaries in industrial, administrative and academic environments. To collect the dataset from the educational institution (schools and universities), we had to get a permission from the head of each department to arrange the time slot and the place for collecting the samples. Then, twenty to thirty minutes were given to the volunteers to complete the form and write their letters. Volunteers were asked to fill up an information page about their gender, age, handedness and work position. Figure 3.3 shows the information page used to collect the writer information for labelling purposes.

The process of labelling the handwritten documents was as follows. First, each sample was given a unique identifier in an Excel sheet ($Doc_1, Doc_2, Doc_3, ...Doc_n$). Second, the writers' information was then extracted from the information page and saved for each document, like in Table 3.1, which shows the labelling of five different handwritten specimens.

	Gender	**Age**	**Handedness**	**Position**	**Label**
Doc1	Female (F)	16	Right (R)	School Student (SS)	F_15_R_SS
Doc2	Male (M)	20	Left (L)	University Student (US)	M_20_L_US
Doc3	Male (M)	30	Right (R)	Public Job (PJ)	M_30_R_PJ
Doc4	Female (F)	35	Right (R)	Private Job (PRJ)	F_35_R_PRJ
Doc5	Female (F)	45	Right (R)	No Job (NJ)	F_45_R_NJ

Table 3.1: Example of labelling five handwritten documents of the FSHS dataset

We have carried a comprehensive manual cleaning process for all the dataset documents. As a result, we excluded the empty samples, files with no information about the writer, and documents with less than five text lines. For example, some of the writers draw ruled lines on the white papers like in figure 3.5, which do not concern our work. The total number of the excluded samples is 272.

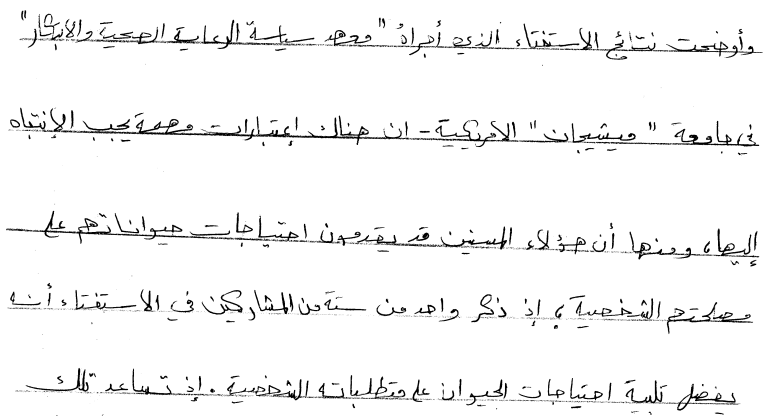


Figure 3.5: Example of a discarded handwritten document from the FSHS dataset

What is significant about our dataset is its size and the number of writers. The variety in age and the large number of writers accounted for the great diversity in the handwriting styles. The dataset contains documents that are mainly written in Arabic; yet, there are still around 15 samples that

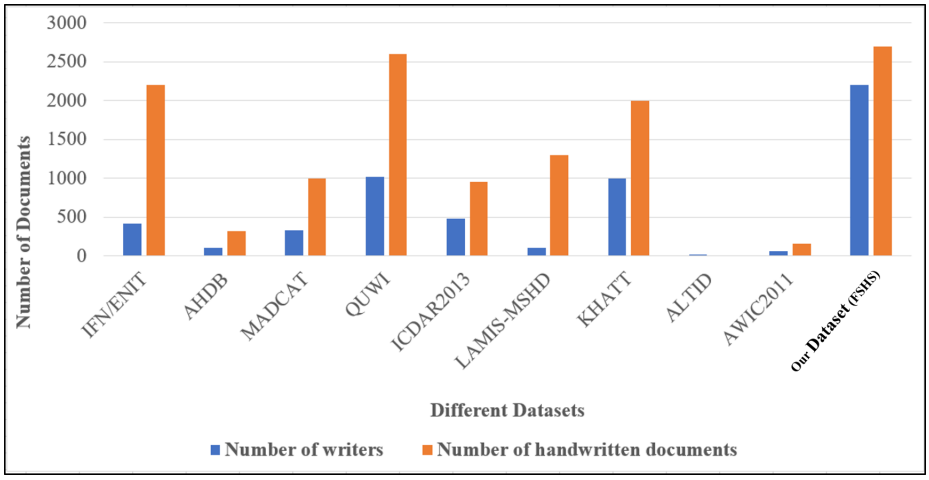


Figure 3.6: Comparison between our dataset and other datasets

were written in English. We digitized the handwritten samples with a resolution of 600 dpi. In total, the dataset contains 2428 digitized pages after cleaning. 43% of the samples were written by males, which equals to 1044 documents, and 57% were written by female writers, which equals to 1383 documents. Each page includes a minimum of five lines with 12140 lines for the whole dataset. Each text line contains approximately ten words yielding 121400 words for the entire dataset.

Figure 3.6 shows a comparison between our dataset (FSHS) and state of the art Arabic datasets. It can be noticed that the FSHS dataset has the highest number of writers and documents among the other available datasets. Table 3.2 shows the statistics of all writers' demographics information. As shown, most writers used the right hand to write, as their percentage was 98.1%. While the highest age were in two ranges, the first 19 and 21 range and the second is over 35 years old. Regarding the occupation, the highest percentage was workers. Figure 3.7 displays the distribution of the FSHS writers in terms of age, gender, occupation and handedness.

The FSHS dataset can be used in many research areas related to human interaction, such as handwriting recognition if appropriately labelled, gender, and age classification.

	Categories	Numbers	Percentage %
Gender	Male	1044	43%
	Female	1383	57%
Age	<= 15	117	4.82%
	16-18	326	13.43%
	19-21	539	22.20%
	22-24	354	14.58%
	25-27	102	4.20%
	28-30	102	4.20%
	31-33	221	9.10%
	34-35	120	4.94%
	>35	547	22.53%
Handedness	Left	46	1.90%
	Right	2382	98.10%
Occupation	School Student (SS)	308	12.69%
	University Student (US)	851	35.05%
	Public Job (PJ) + Private Job (PRJ)	1239	51.03%
	No Job (NJ)	30	1.24%

Table 3.2: Statistics of the number of the FSHS dataset writers in terms of gender, age, handedness and occupation

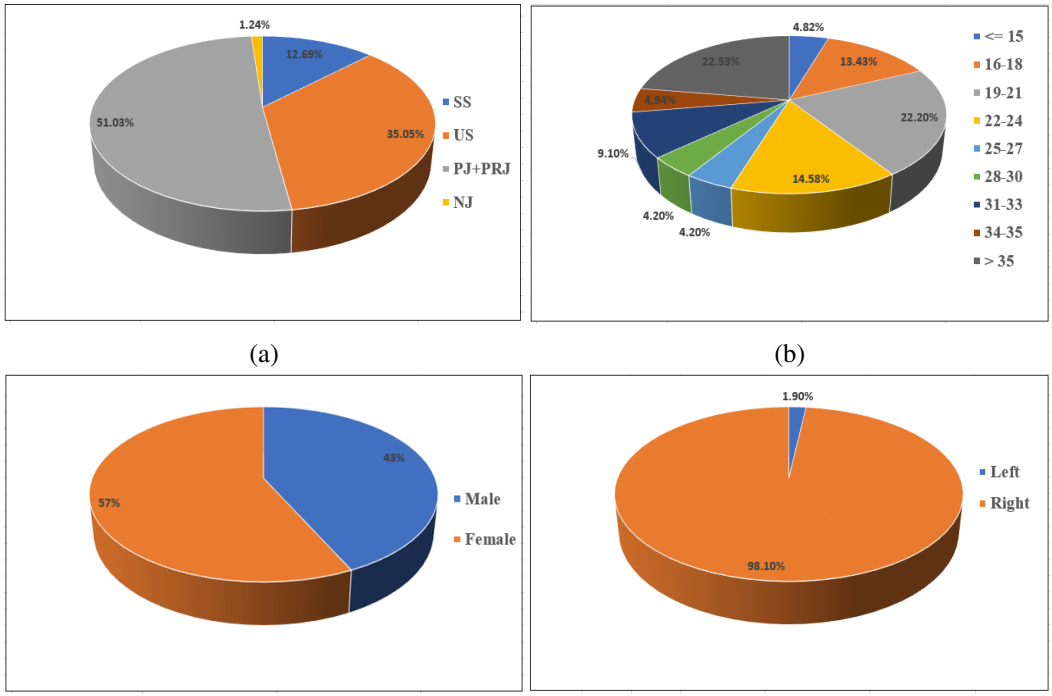


Figure 3.7: Distribution of writers with respect to (a) Occupation, (b) Age, (c) Gender and (d) Handedness

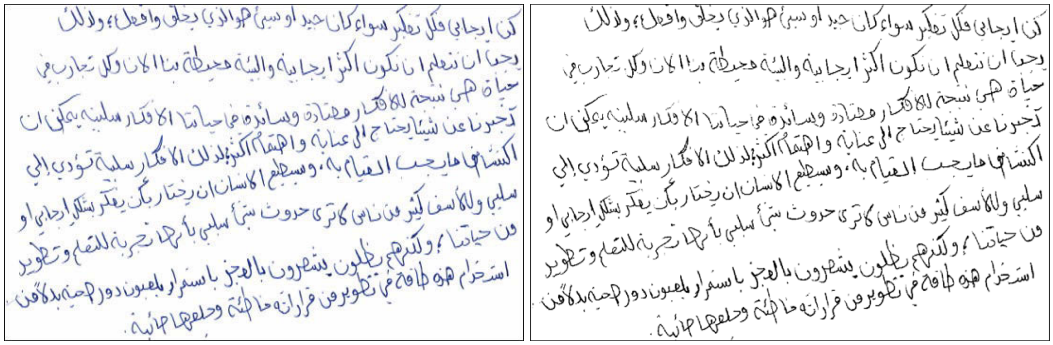
3.2 Preprocessing

The preprocessing step is required to improve the quality of images before the feature extraction phase. In handwriting analysis systems, many preprocessing techniques can be applied to improve the quality of handwriting documents, such as binarization, noise removal, skew angle correction, thinning and skeletonization. In our research, it is necessary to preserve the main features of the handwriting because it helps to differentiate the gender and the age of the writer, so we applied simple preprocessing techniques, which is limited to noise removal using a 3×3 median filter and binarization using OTSU thresholding [73]. Median filter reduces the noise in images but preserves the valuable details.

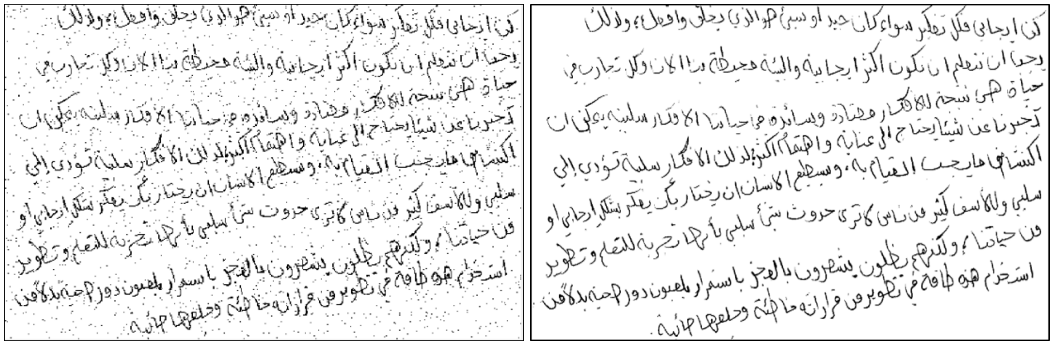
Figure 3.8 shows the preprocessing results in a handwritten sample, where (a) is the coloured original sample and (b) is the binarized version. On the other hand, (c) is a handwritten document with salt and pepper noise removed by applying a 3x3 filter, as shown in (d), which is filtered and has no noise.

A cropping of the handwriting was applied by first finding the first black pixels in each of the four directions (Left_limit, right_limit, upper_limit, lower_limit), and then the writing was separated image [lower_limit: upper_limit, Left_limit:right_limit]. Figure 3.9 shows an example of the cropping process where the green circles show the first hit of the black pixels from all direction.

On the other hand, some of the extracted features in our work are sensitive to noise, such as handwriting irregularity and word spacing, so we applied noise removal then we used dilation [74]. Where, text document T is dilated using structure element S to produce a dilated document D as in equation 1. Dilation in image processing is one of the morphological operations. It adds pixels to the boundaries of an objects in a binary image using a matrix of 1's and 0's called structuring element. Dilation is used to bridge gaps in an image. As shown in Figure 3.21, a handwritten



(a) Colored image (b) Binarized image



(c) Noisy image (d) Filtered image

Figure 3.8: Preeprocessing in terms of binarization and filtering

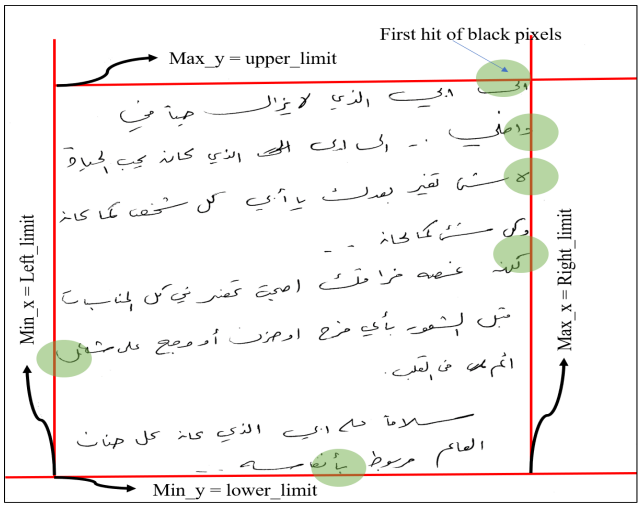
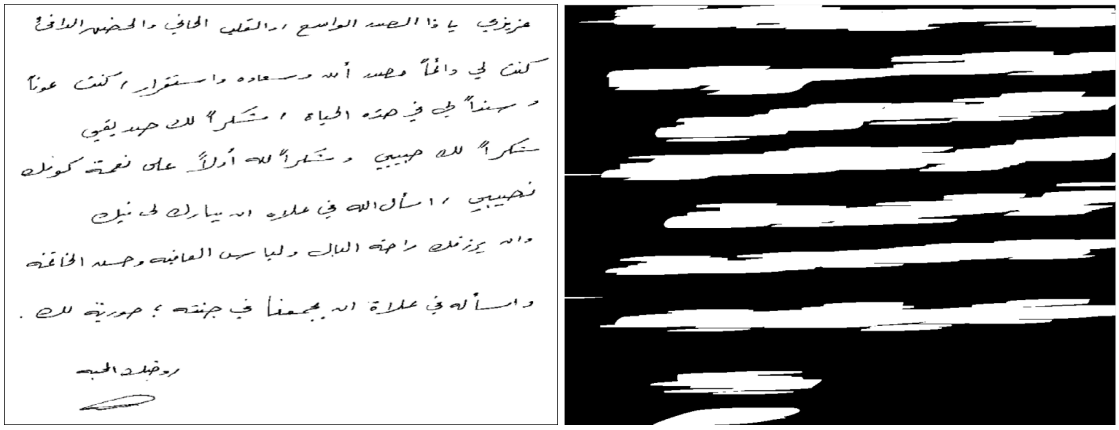


Figure 3.9: The cropping of handwriting

document in (a) and its corresponding blobs in (b).

$$D = T \oplus S = \{ z \mid [(\hat{B}_z) \cap T] \subseteq T \} \tag{1}$$

Where $\hat{S} = \{\mathrm{c} \mid \mathrm{c} = \text{-a}, \mathrm{a} \in S\}$, and D, T, S are sets in Z^2.



(a) Original image (b) Dilated image

Figure 3.10: Example of dilated handwriting document

3.3 Feature Extraction

In our work, we extracted two different types of features. The first one uses image processing manipulation functions to process the images and extract the desired features such as pen pressure, word spacing and irregularity. The second method utilizes pre-trained CNN networks to extract the features automatically from the handwriting documents.

3.3.1 Feature Extraction Using Image Processing Techniques

Gender Related Features:

A set of features were extracted from Arabic handwritten documents to detect the writers' gender. A psychologist and a graphologist also recommended the selected features. These features are pen pressure (PP) with a vector of 2 features, irregularity in text line (TLI) with a vector of 700 features, the word spacing (WS) with 30 features and the margins (LRM) with 400 features. We examined these features separately and combined by concatenating them. For example, one of the

combinations is PP (2 features) + TLI (700 features).

We studied different lengths of each feature to find its effect on the performance of the detection system. SVM method was applied on 2000 (15% for validation, 15% for testing and 70% for training) handwritten samples from the FSHS dataset to choose the optimal length of each feature. A full description of each feature will be explained in the following section.

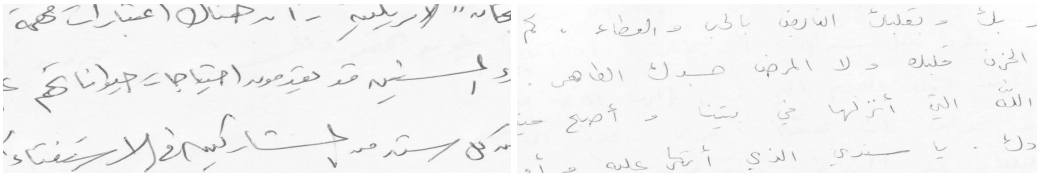
Pen pressure, most of the writers used a dry blue pen to write their letter, which make the extraction of this feature more precise and accurate. As male writing in general shows heavier pen pressure than females. Figure 3.12 shows two different handwriting specimens, where (a) is written by a male writer, while (b) is written by a female writer. It is obvious that the writing in (a) is heavier than the writing in (b). We calculated the standard deviation (SD) and mean X' , to extract the pen pressure feature from the gray-scale handwritten documents using equations 2 and 3 respectively.

$$SD = \sqrt{\frac{1}{N-1} \times \sum_{i=1}^{n} \sum_{j=1}^{m} (X_{(i,j)} - X')^2} \tag{2}$$

$$X' = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (X_{(i,j)})}{N} \tag{3}$$

where, X' is the mean value of the pixel intensities of the image, N is the number of pixels in an image $(n \times m)$, and $X_{(i,j)}$ is the intensity value of each pixel.

We have calculated the mean and the standard deviation of all the samples we collected, and we found that the light pressure writing (female) mostly has a standard deviation higher than the heavy pressure writing (male). This can be shown in Figure 3.12, which illustrates the distribution of the SD values for a set of 120 handwriting samples, and it can be noticed that female handwritings have a higher standard deviation. This result is to be expected due to the physical characteristics of males such as his strength, height and larger hands.



(a) Male heavy writing (b) Female light writing

Figure 3.11: Difference between female and male pen pressure

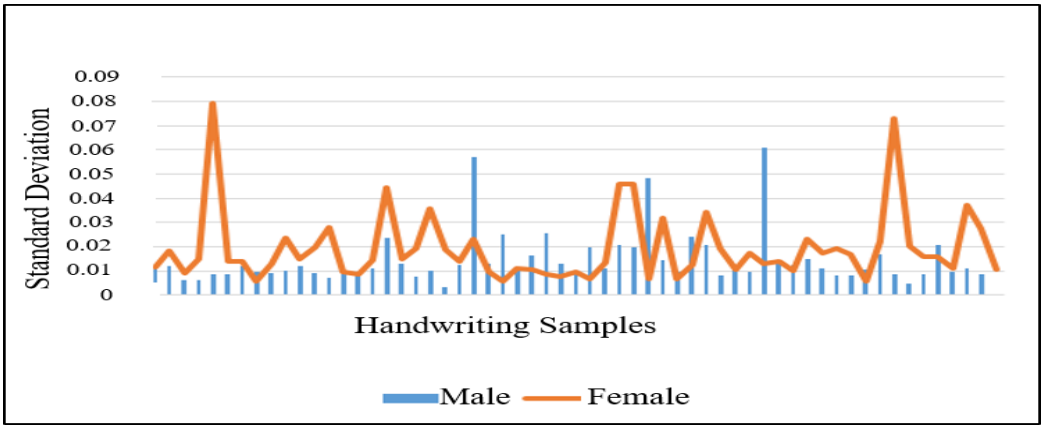


Figure 3.12: SD Distribution of a set of handwriting samples

The space between words, according to graphologists, the space between words gives an indication on the writer's gender. In which, females tend to have narrow spacing between words, while men have an average larger spacing.

Arabic writing is a horizontal cursive script from right to left whose words consist of sub-words also called Pieces of Arabic Words, each of which consists of one or more letters, this can be shown in Figure 3.13. Sub-words are formed due to the fact that there are some Arabic letters that cannot be connected to any other letter, causing a discontinuity in the cursiveness of the word. Consequently, there is no difference between the within and between word spacing in Arabic writing.

Figure 3.13: Arabic handwritten text-line

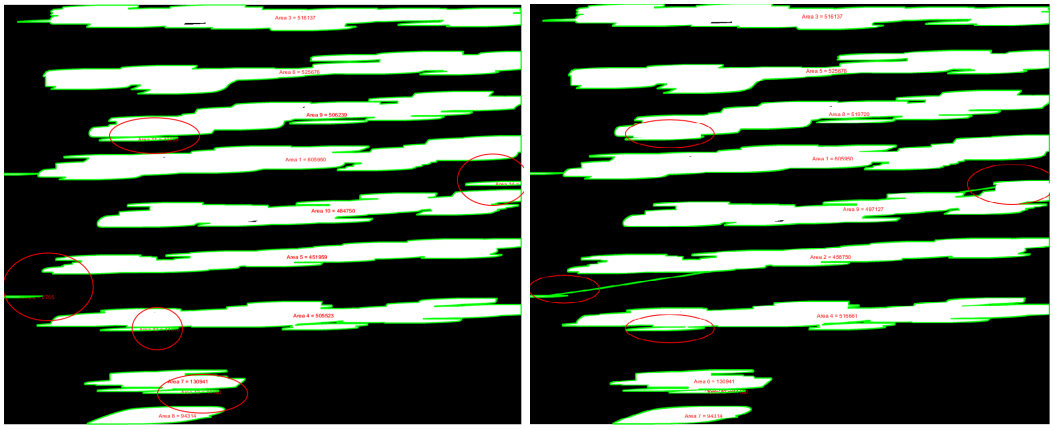
As a result, we chose to measure the distance between sub-words instead of words for Arabic writing. Word spacing features were extracted from the document, by first extracting the text-line from the document [74]. Then, the sub-words were extracted by the connected components of each line [75]. Projection profiles have been widely used in text segmentation. In our research, we used the projection profiles to analyse the text document and extract some of its properties such as, the peaks and valleys of the projection profile as well as the slope of each text line. After analysing the document, all the connected components are found, where the vertical projection profile f $(p(x), x)$ is found to calculate the average width W_{avtb} of the text line. While the horizontal profile $g(y, p(y))$ is also calculated to find the height H_{avtb} of the text line. Then, the slope between the peaks and valleys of the projection profiles is found to determine the spaces between the different text-lines. The slope ($slope_h$) is calculated using the following equation:

$$Slope_h = \frac{P(y_{peak}) - P(y_{valley})}{(y_{peak} - y_{valley})}, \tag{4}$$

where, $p(y)$ is the number of white pixels in row y, while y_{peak} and y_{valley} are the coordinates of the rows where the peaks and valleys are located.

One of the most important characteristics of Arabic scripts is diacritics and dots. These are small connected components that are located above or below the words, as you can see in Figure 3.13. Thus, they are often located between the text lines which makes the segmentation process harder. The segmentation algorithm then starts by replacing an adaptive mask m around each text-line blob that has the height and the width within the blob area and merging the ones that are very close to the blob. Figure 3.14 (a) shows the original blobs of all connected components of the text, while Figure 3.14 (b) shows the merged blobs with the closest ones.

Connected components [75], in a 2D image, are clusters of pixels with the same value, which are connected to each other through either 4-pixel, or 8-pixel connectivity. In our experiments we used 8-connectivity where 8-connected pixels are neighbours to every pixel that touches one of their edges or corners. These pixels are connected horizontally, vertically, and diagonally. A set S



(a) Blobs of document (b) Merged blobs

Figure 3.14: Diacritics removing process in text segmentation

of pixels form a connected component if there is at least one path in S that joins every pair p , q of pixels in S, the path must contain only pixels in S [76]. After the segmentation of the text-lines, the distance between the connected components is calculated by finding the distance between the horizontal location of the last black pixel from one word with the first black pixel of the following word using Euclidean distance method [77]. As, if $p = (p_1, p_2)$ and $q = (q_1, q_2)$, the distance is given by the following equation.

$$D(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \tag{5}$$

The extracted feature was generated by creating a histogram of 30 experimentally calculated bins. Table 3.3 shows different accuracy rates obtained using various bin sizes. Based on these results, we can see that 30 bins are sufficient to get a reasonable classification rate, so it was chosen to find the word spacing features. The results of different bins are visualized in Figure 3.15. Each bin represents the number of distances between words within a given range. For example, the first bin is the number of times the distance between words is between 1 and 5 pixels. The second bin is a count of the number of times the distance between the words within the document is between 6 and 10, and so on.

Word Spacing	
Number of bins	**Accuracy (100%)**
4	72.8
6	72
8	73
10	73.8
12	72.5
14	73.6
16	71.2
18	72.8
20	74.1
22	73.6
24	73.4
26	73.9
28	74.5
30	74.9

Table 3.3: Accuracy rates for different word spacing bin sizes

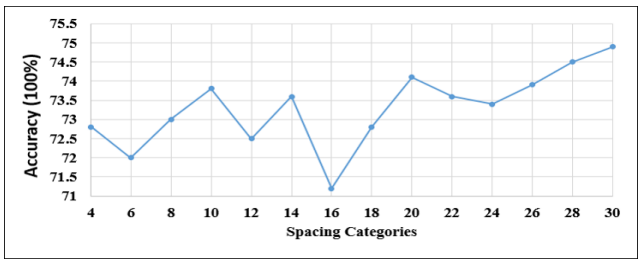
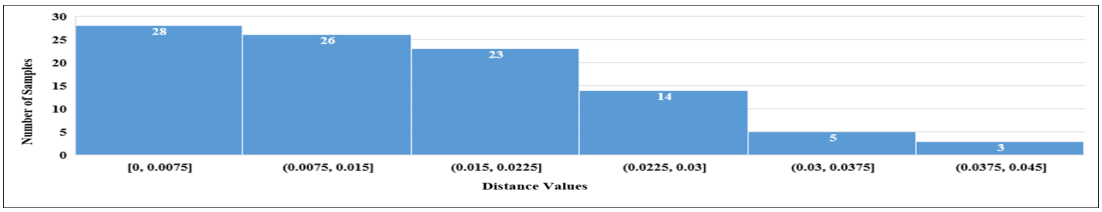
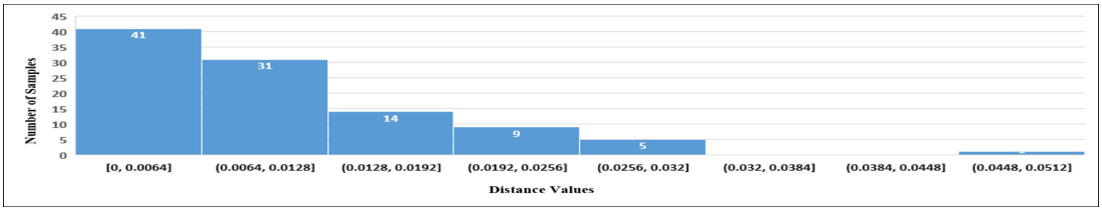


Figure 3.15: Comparison between different accuracy rates of different word spacing bins



(a) Male handwriting samples distribution

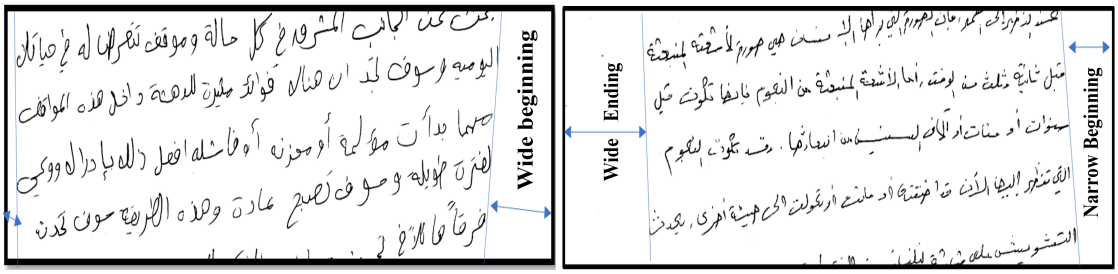


(b) Female handwriting samples distribution

Figure 3.16: The Difference between female and male handwriting in terms of word spaces

Figure 3.16 shows the differences between female and male handwriting in terms of word spacing. It can be noticed that females tend to have smaller spaces between words, while males are not consistent with the spacing. Males may have medium and large spacing between words or subwords in the case of Arabic writing.

Left and right margins are considered one of the most important features to distinguish between female and male, as female is recognised by a narrow beginning margin and a wide ending one. However, male writing is the opposite, as they tend to have wide beginning of a line and narrow ending. This can be shown in Figure 3.17.



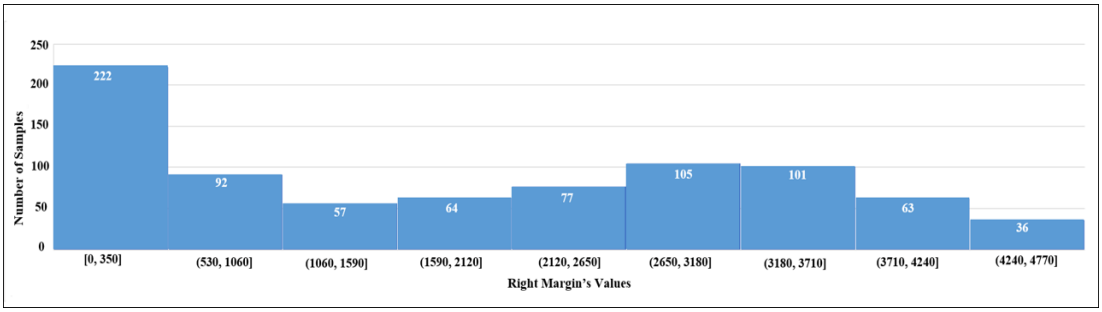
(a) Margins in male handwriting (b) Margins in female handwriting

Figure 3.17: Margin difference between female and male handwriting

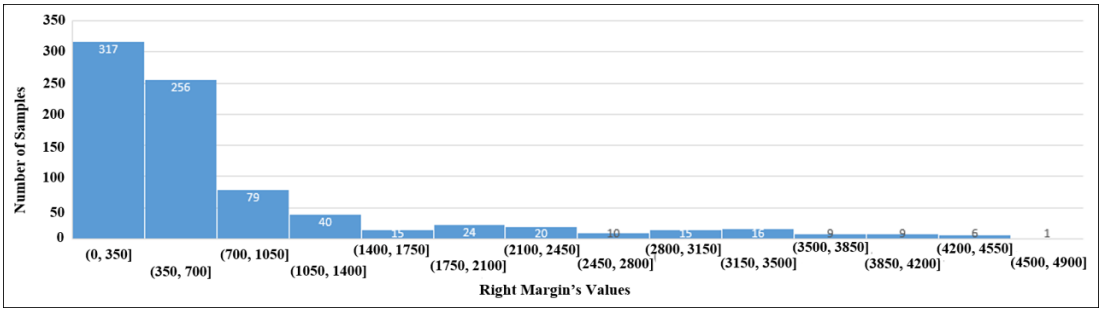
Figures 3.18 and 3.19 also show the distribution of male and female samples in terms of left and right margins. Figure 3.18 (b) which is written by female writers, show that only a few samples have large right margins. While in (a) great amount of samples have a large right margin, in which those samples are written by male writers.

Figure 3.19 shows the left margin distribution in male and female samples. Figure (a) which illustrates the left margin of documents written by female writers, a significantly large number of documents have large left margin, while in (b) where the documents are written by male writers, most of the documents have narrow left margin.

To measure the left and right margins of the handwritten documents, the documents were first dilated using equation 1. Then, the left margins (end of the line) R_m were found by finding the

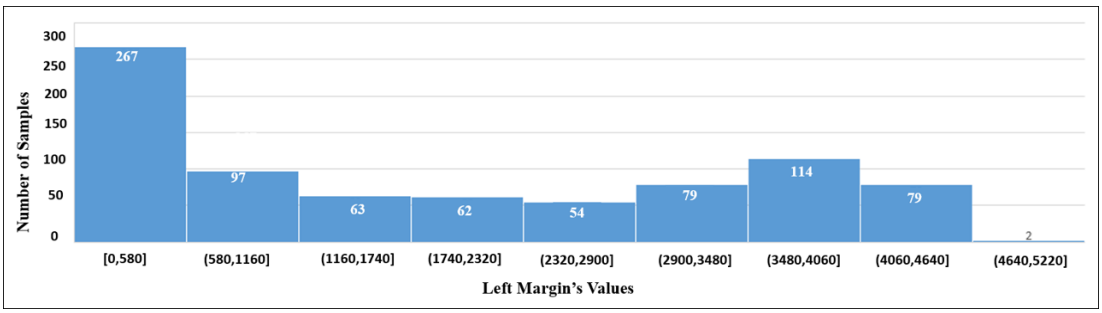


(a) Right margin in male handwriting

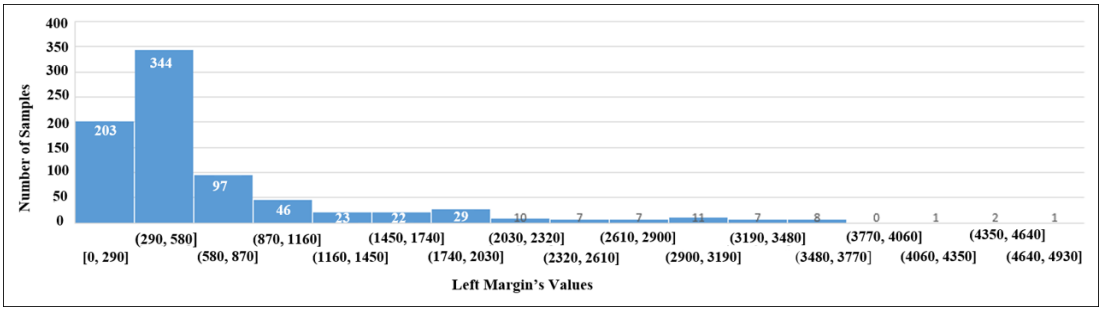


(b) Right margin in female handwriting

Figure 3.18: Right margin feature in female and male handwriting



(a) Left margin in female handwriting



(b) Left margin in male handwriting

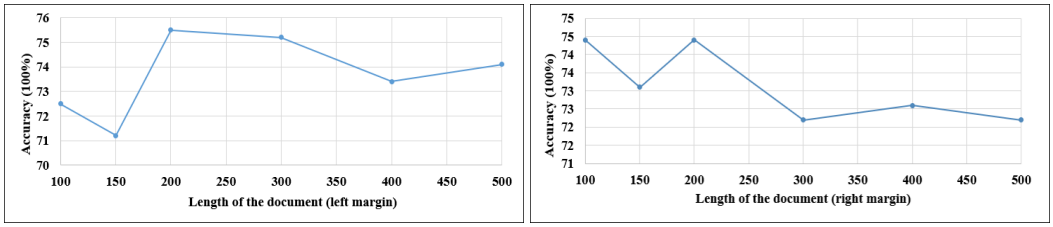
Figure 3.19: Left margin feature in female and male handwriting

hits of the first black pixels from the left. We had experimentally discovered that finding the first pixels of the only first 200 rows of the image document is sufficient to measure the margin, and this can be shown in Table 3.4 and Figure 3.20 where the accuracy rate was the highest when we used 200 rows for each of the left and the right margins. Thus, if the first 200 black pixel coordinates are (x_1, y_1), (x_2, y_2), (x_3, y_3) ..... (x_{200}, y_{200}), where x_i is the first hit pixel and y_i is the row number (y_i is the ith row, and y_7 is the 7th row). Then, the resulting left margin (end of the line) feature will be (x_1, x_2, x_3, ..., x_{200}).

For the right margin (beginning of the line) the last black pixel of the line (last hit) is subtracted from the width of the document W_{doc}. Similar to the left margin features, the right margin features are calculated as follows: for a set of points (x_1, y_1), (x_2, y_2), (x_3, y_3) ..... (x_{200}, y_{200}), where x_i is the last pixel of the line and y_i is the line number, then the extracted feature vector will be ($W_{doc} - x_1$, $W_{doc} - x_2$, $W_{doc} - x_3$, ..., $W_{doc} - x_{200}$), similar to the left Margin, 200 rows were experimentally calculated to be sufficient to extract the right margin feature vector.

Left Margin		**Right Margin**	
Length (pixels)	**Accuracy (100%)**	**Length (pixels)**	**Accuracy (100%)**
50	74.2	50	73.8
100	72.5	100	74.4
150	71.2	150	73.1
200	75.5	200	74.4
300	75.2	300	72.2
400	73.4	400	72.6
500	74.1	500	72.2

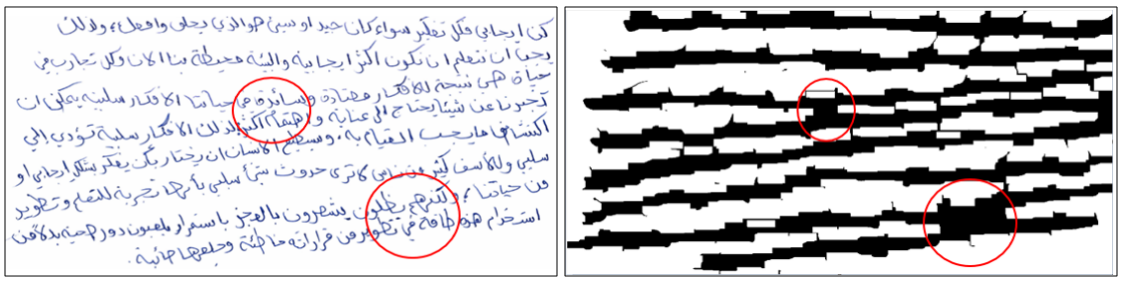
Table 3.4: Different accuracies for different margin lengths



(a) Left margin (b) Right margin

Figure 3.20: Comparison of different accuracy rates for different margin lengths

Handwriting irregularity, there are many aspects of irregularity in handwriting, such as: slant irregularity, and the irregularity between handwritten lines. We focused on extracting irregularity between handwritten lines. We proposed measuring lines irregularity by applying the horizontal and vertical projection profiles after dilating the text in a given document using equation 1. We used dilation before the extraction of some features such as margins, handwritten irregularity and word spacing. Figure 3.21 shows that the overlapping between text lines in (a) is easy to be discovered and recognised in the dilated image in (b).



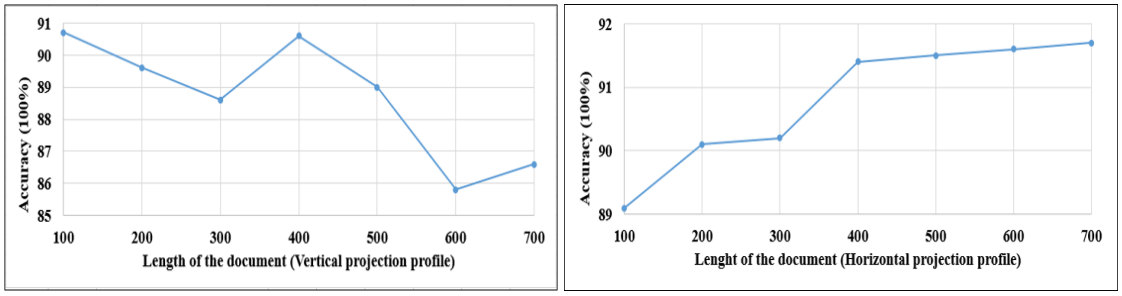
(a) Original image (b) Dilated image

Figure 3.21: Overlapping between text-lines

The horizontal projection profile f(y, p(y)) and the vertical projection profile f(x, p(x)) of the document are found for the first n rows of the document, which was experimentally chosen to be the first 700 rows for each projection. Table 3.5 shows the different experiments conducted to find the optimal length of the projection profiles. It can be seen that the length of 700 pixels for the horizontal profile (H) gave the highest accuracy. However, for the vertical projection (V), the one with 100 pixels has the best performance. To choose the optimal lengths for both projection profiles, we tested three different length combinations. The combinations are 700 pixels for the vertical profile with 700 pixels for the horizontal profile, 100 pixels for a vertical profile with 700 pixels for a horizontal profile and 400 pixels for the vertical profile with 700 pixels for the horizontal profile. Figure 3.23 shows the results of each combination. As the highest accuracy was obtained using 700 pixels for each projection profile. Figure 3.22 visualised the results obtained from different vertical projection profile lengths in (a) and for horizontal projection profile in (b) .

Vertical projection profile		Horizontal projection profile	
Length (pixels)	**Accuracy (100%)**	**Length (pixels)**	**Accuracy (100%)**
100	90.7	100	89.1
200	89.6	200	90.1
300	88.6	300	90.2
400	90.6	400	91.4
500	89.0	500	91.5
600	85.8	600	91.6
700	86.6	700	91.7

Table 3.5: Different accuracy rates of different vertical and horizontal projection profiles



(a) (b)

Figure 3.22: Accuracy rates of different projection profiles where (a) vertical projection and (b) horizontal projection

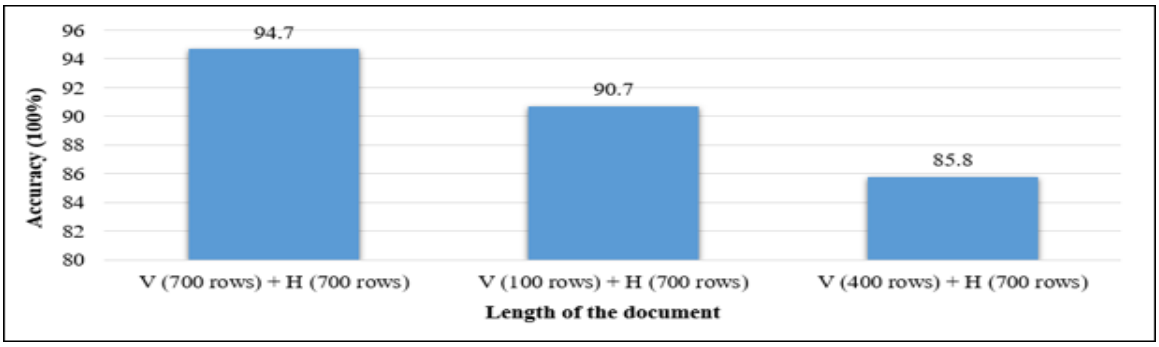
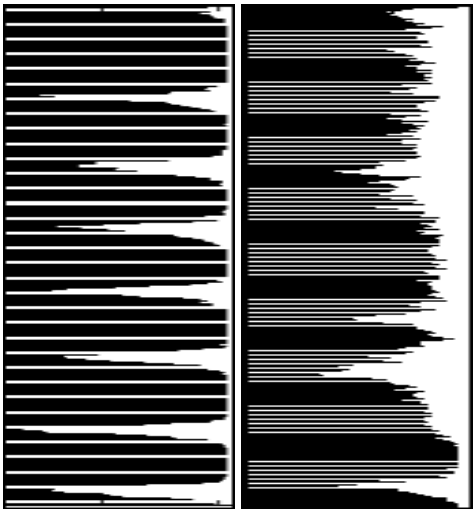


Figure 3.23: Different combinations of horizontal and vertical projection profiles

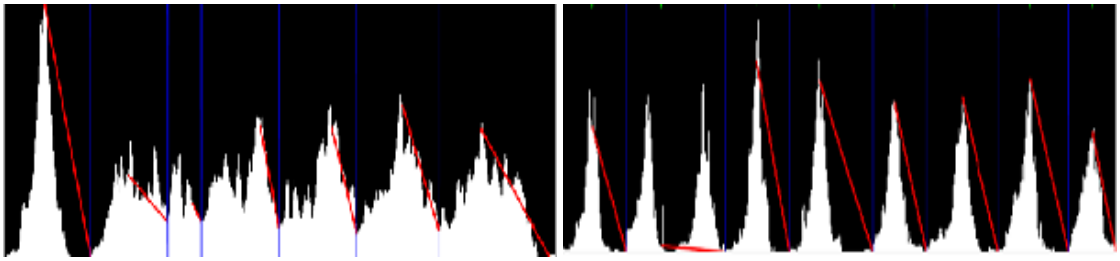
The projection profile reflects the nature of the document and the distribution of the text lines. The algorithms consider the significant peaks and valleys.

Figure 3.25 shows the horizontal projection profile of two different documents. The horizontal projection profile in Figure 3.25 (a) shows that the lines are irregular and not well separated. In addition, there are no deep valleys in many parts of the profile, which gives an indication that the



(a) (b)

Figure 3.24: Vertical projection profile of two different documents



(a) (b)

Figure 3.25: Horizontal projection profile of two different documents

lines are skewed and close to each other. The profile of Figure 3.25 (b) shows that the lines in the documents are more regular and are nicely separated, since it has deep valleys and peaks.

Figure 3.24 shows the vertical projection profile of two different documents. Figure 3.24 (a) reflects the nature of regular document, while in (b), it shows irregular documents which have no uniform peaks and valleys

3.3.2 Age Related Features

A set of features were extracted from Arabic handwritten documents to detect the writer's age. A psychologist and a graphologist also recommended the selected features. These features are Irregularity in pen pressure (IPP) with a vector of 200 features, irregularity in slant (IS) with a feature

vector of 100 features, irregularity in text line (TLI) with a vector of 700 features, and the percentage of white and black pixels (PWB) with 4 features. We examined these features separately and combined by concatenating them. For example, one of the combinations is IPP (200 features) + IS (100 features).

To find the optimal length of each age-related feature, we applied the SVM method on 2000 (15% for validation, 15% for testing and 70% for training) handwritten samples of the FSHS dataset, then the length with the highest accuracy was chosen. A full description of each feature will be explained in the following section.

Slant irregularity: As mentioned before, Arabic writing is a horizontal cursive script whose words consist of sub-words or Pieces of Arabic Words (PAWs), each consisting of one or more letters; this can be shown in Figure 3.13 on page 40, where each colour corresponds to a single separated CC. Consequently, a connected component (CC) [75] (sub-word) can be used to find the irregularity in slant.

To find the irregularity in slant, we calculated the CCs of the binary images. Then, we applied the method of least square [78] to find the slant by locating the best fit line of each CC. Given a set of n points, the best fit can be found using equations 6 and to find the coefficients a and b, equations 7 and 8 were applied, respectively, while equation 9 was used to find the skew angle (α). The number of CCs used to find this feature was empirically set to 100 for FSHS dataset and 200 for Khatt dataset, with $size(CC) > 10$ pixels.

$$y = a + bx \tag{6}$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{7}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} \tag{8}$$

$$\alpha = \tan^{-1}(b) \tag{9}$$

SVM method was utilized to conduct several experiments using different groups of CCs to determine the best one. The experiments were divided into three phases. In the first stage, the CCs groups were tested to detect the writer's age on female handwriting. Table 3.6 shows different accuracy rates obtained using CCs groups ranging from 50 to 250. It can be noticed that most of the results are close, but despite that, the 100 CCs group gave the highest performance of 56.5% in accuracy. These results are visualized in Figure 3.26.

Slant Irregularity (Female)	
Number of CCs	**Accuracy rate (100%)**
50	53.2
100	56.5
150	54.5
200	51.6
250	51.6

Table 3.6: Results of using different number of CCs on female handwriting

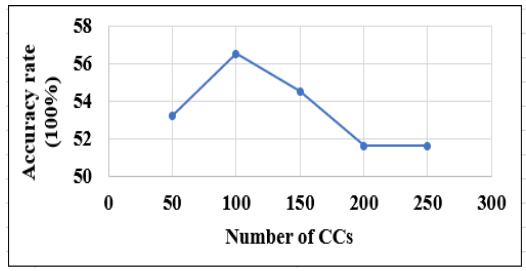


Figure 3.26: Corresponding accuracy rates using different square numbers

Slant Irregularity (Male)	
Number of CCs	**Accuracy rate (100%)**
50	56
100	57.2
150	58.8
200	59.5
250	62.9

Table 3.7: Results of using different number of CCs on male handwriting

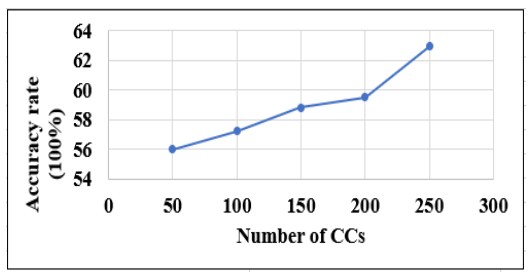


Figure 3.27: Corresponding accuracy rates using different square numbers

On the other hand, we conducted the same experiments and used the same CCs groups, but on handwriting for males only. As a result, we obtained the accuracy rates shown in Table 3.7. The highest performance occurred when we used the set of 250 CCs and got 62.9%. These results are visualized in Figure 3.27.

Slant Irregularity (Female & Male)	
Number of CCs	**Accuracy rate (100%)**
50	50.2
100	56.3
150	53.5
200	54.2
250	54.3

Table 3.8: Results of using different number of CCs on female and male handwriting

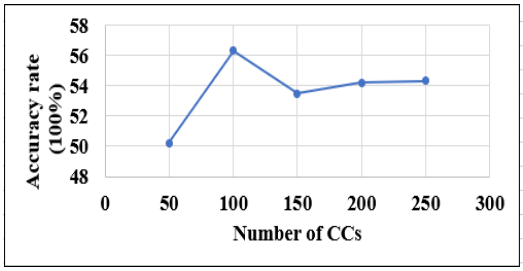


Figure 3.28: Corresponding accuracy rate of different number of CCs

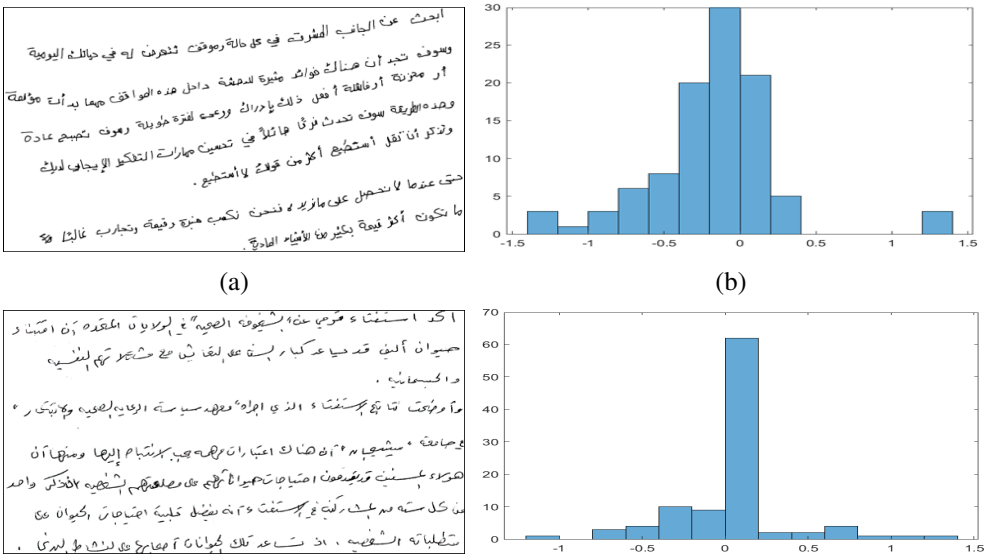


Figure 3.29: Handwriting samples with their corresponding slant irregularity distribution

For both female and male handwritings, the experiments showed that the best group to detect writer's age is 100 CCs. The results of different groups can be seen in Table 3.8, where the highest performance rate is 56.3% when using the 100 CCs group. These results are visualized in Figure 3.28.

Based on the previous results of slant irregularity experiments, we found that youth adult's handwritings are more irregular than mature adults. For example, Figure 3.29 (a) shows the handwriting of a 19 years old female and its corresponding slant distribution in (b), where it is so obvious that the slants are irregular and distributed over different angles. On the other hand, in (c), which is the handwriting of 44 years old female, the handwriting is more regular, which can be seen in (d).

Pen pressure irregularity: to find the irregularity in pen pressure, the grey-scale image was divided into uniform squares of size 100×100 pixels. The standard deviation (SD) and mean X' were calculated for each square separately, using equations 10 and 11, respectively. The number of the segmented squares was empirically set to 200.

$$SD = \sqrt{\frac{1}{N-1} \times \sum_{i=1}^{n} \sum_{j=1}^{m} (X_{(i,j)} - X')^2} \tag{10}$$

$$X' = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (X_{(i,j)})}{N} \tag{11}$$

where, X' is the mean value of the pixel intensities of the square, N is the number of pixels in a square $(n \times m)$, and $X_{(i,j)}$ is the intensity value of each pixel.

Many experiments were conducted using the SVM method to determine the optimal number of squares to extract the pen pressure irregularity for age detection from handwritten documents. The experiments were divided into three phases. In the first stage, the different groups of squares were tested on female handwriting. Table 3.9 shows different accuracy rates obtained using square groups ranging from 50 to 200, where the 100 uniform Squares group gave the highest performance by 63.6% of accuracy. These results are visualized in Figure 3.30.

Pen Pressure Irregularity (Female)	
Number of squares	**Accuracy rate (100%)**
50	53.4
100	63.6
150	58
200	52.2

Table 3.9: Different number of square results on female handwriting

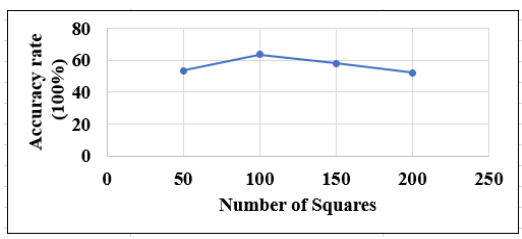


Figure 3.30: Corresponding accuracy rates using different square numbers

Moreover, testing the same square ranges on the handwriting of males only gave the results shown in Table 3.10, where the 100 square groups gave the highest accuracy of 54.5%. These results are shown in Figure 3.31.

Pen Pressure Irregularity (Male)	
Number of squares	**Accuracy rate (100%)**
50	52.5
100	54.5
150	52.1
200	53.7

Table 3.10: Different number of square results on male handwriting

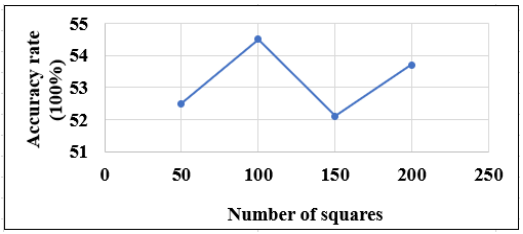


Figure 3.31: Corresponding accuracy rates using different square numbers

Pen Pressure Irregularity (Female & Male)	
Number of squares	**Accuracy rate (100%)**
50	52.5
100	53.7
150	52.3
200	53.8

Table 3.11: Different number of square results on female and male handwriting

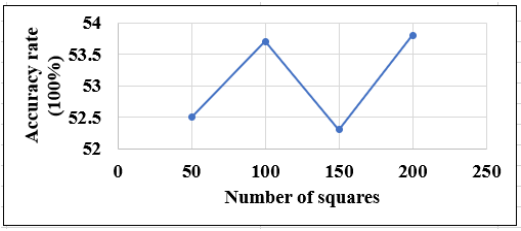


Figure 3.32: Corresponding accuracy rates using different square numbers

Per contra, the experiments showed that the best square group to extract the pen pressure irregularity is 200 squares for both female and male handwritings. Table 3.11 shows different accuracy rates for different square groups. The highest performance rate is 53.8% when using the 100 squares. These results are shown in Figure 3.32.

Figure 3.33 shows two different handwritings and their corresponding pen pressure distributions where (a) belongs to a youth-adult writer, with an irregular pen pressure distribution in (b). In contrast, (c) is a mature adult's handwriting, and (d) is its corresponding regular pen pressure distribution.

Percentage of black and white pixels: in a handwritten document, this feature is an indication of the width of the handwriting. This process starts by cropping the image to a square with a width of m pixels. The size of the square was empirically chosen to be of size 250 x 250 pixels for the FSHS dataset, and 100 x 100 pixels for the khatt dataset. The ratio of the black pixels was calculated as follows: the number of black pixels in the cropped square was first calculated. Then the ratio

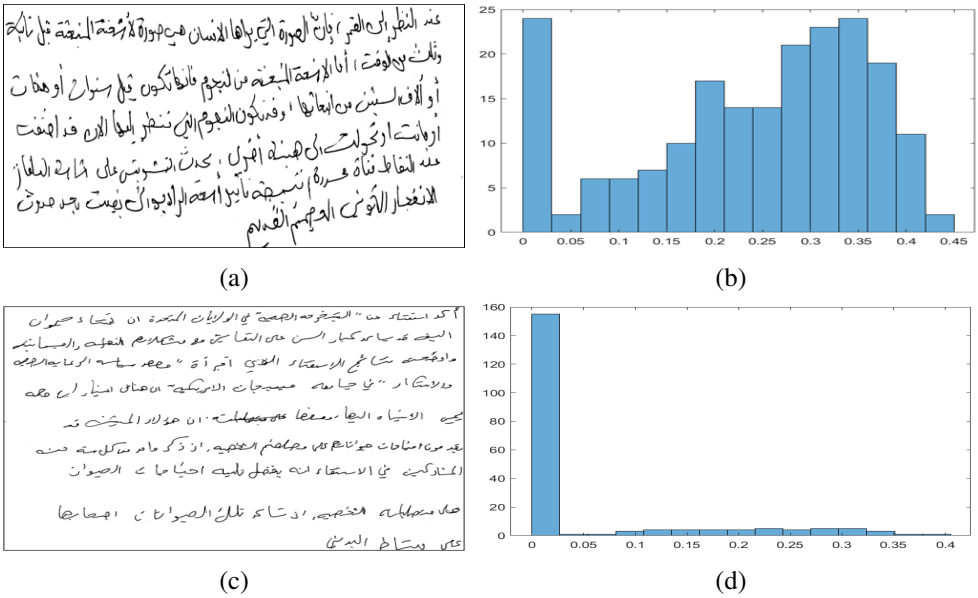


Figure 3.33: Handwriting samples with their corresponding pen pressure irregularity distribution

between the number of black pixels and the size of the square was measured.

This feature consists of a vector of four values: the number of black pixels (BP), the number of white pixels (WP = mxm - BP), the percentage of black pixels (PBP = BP/(mxm)) and white pixels (PWP = WP/(mxm)) of the cropped square.

We examined different sizes ranging from 50 to 250 to find the best square size. Table 3.34 shows the results of applying the SVM method to the handwriting of female writers. It can be seen, the best square size is 150, with an accuracy rate of 53.6%. The visualization of these results can be shown in Figure 3.34.

Table 3.13 shows the results obtained by applying age classification using different square sizes on male handwriting. The produced accuracy rates show that the best square size is 250, where it got an accuracy of 54.1%, which also can be seen in figure 3.35.

Black and white pixels percentage (Female)	
Dimension of the square	**Accuracy rate (100%)**
50	50.4
100	51.9
150	53.6
200	49.3
250	50.7

Table 3.12: Different square dimension results on female handwriting

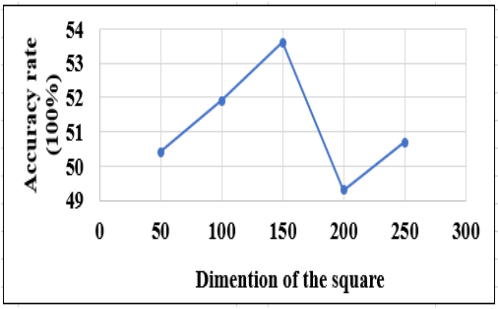


Figure 3.34: Corresponding accuracy rates using different square dimensions

Black and white pixels percentage (male)	
Dimension of the square	**Accuracy rate (100%)**
50	52.5
100	52.5
150	54.
200	53.2
250	54.1

Table 3.13: Different square dimension results on male handwriting

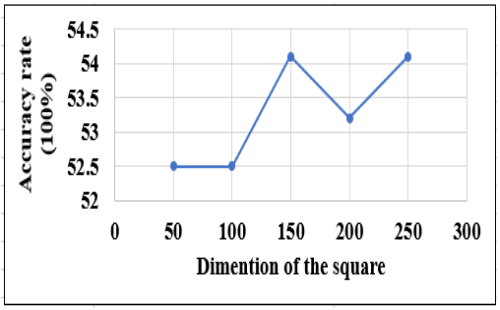


Figure 3.35: Corresponding accuracy rates using different square dimensions

Black and white pixels percentage (Female & Male)	
Dimension of the square	**Accuracy rate (100%)**
50	51.3
100	51.5
150	53.2
200	54.5
250	55.7

Table 3.14: Different square dimension results on Female and male handwriting

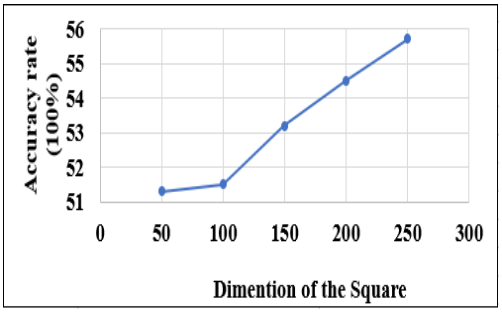
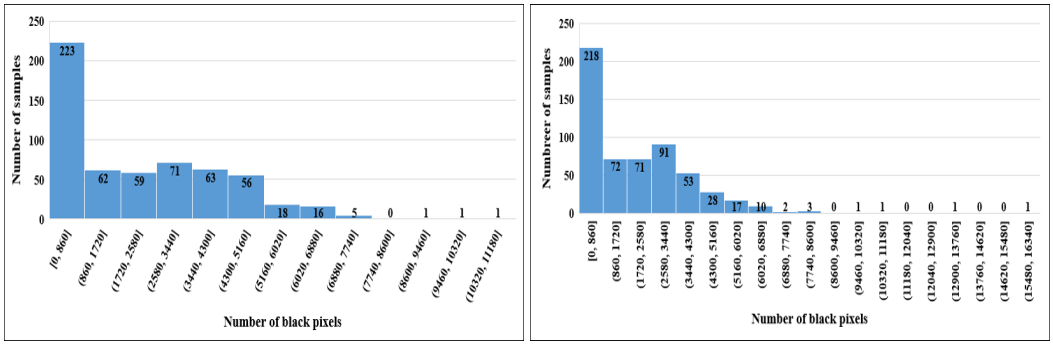


Figure 3.36: Corresponding accuracy rates using different square dimensions

Table 3.14 shows the results of using different square sizes on both female and male handwriting. The square size of 250 got the highest accuracy rate of 55.7%, which is visualised in Figure 3.36.

The percentage of white and black pixels can give us an indication of writing speed. The writer with a higher speed usually has lower black pixels than slower writers. Based on the results we



(a) Youth adult (b) Mature adult

Figure 3.37: Distribution of the number of black pixels in youth and mature handwriting

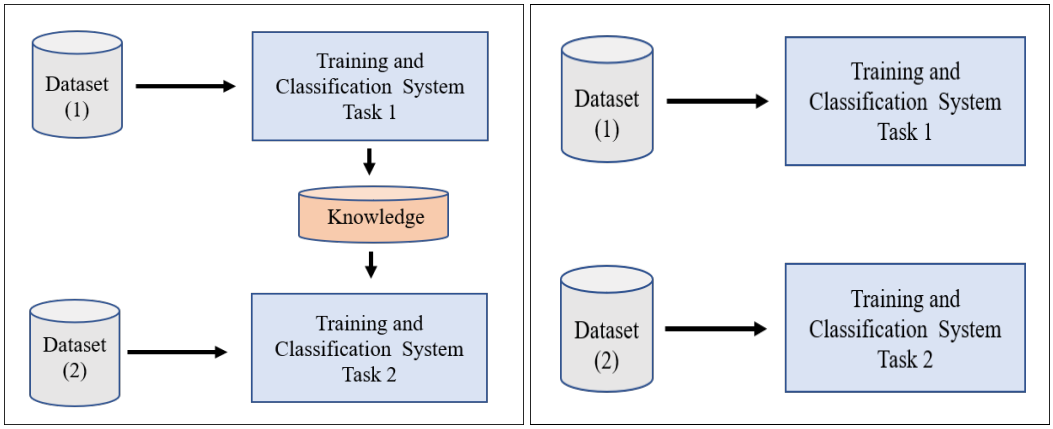
got using this feature and despite the high similarity of the accuracy rates, we noticed that youth people have a less expansive white space than adults, which is consistent with the opinion of the graphologists that youth adults are slower in writing than mature adults.

Figure 3.37 shows the number of black pixels in youth and mature handwriting. It can be noticed that the youth writers relatively have wider writing than mature writers; because of that, they are considered slow writers.

In addition to the previous features, we have used the irregularity between text lines, which was used in gender detection experiments, to reveal the writer's age. It is measured by applying the horizontal projection profile after dilating the text in a given document. The document's horizontal projection profile f (y, p(y)) is found for the first n rows of the document, which was experimentally chosen to be the first 850 pixels for age detection.

Automatic Feature Extraction Using Pre-trained Deep Neural Networks

Convolution Neural Network (CNN) becomes a potent tool for image classification tasks because it automatically extracts the features from raw data without the need for traditional machine learning techniques. This makes the feature extraction process easier and more effective. On the other hand, CNN does not produce a good performance when the dataset is small, where the size of the data must be large to get good results [79]. Collecting a large dataset of handwriting is not an



(a) Transfer learning (b) Machine learning

Figure 3.38: Traditional machine learning vs transfer learning

easy task, so to solve this problem, the concept of transfer-learning was introduced, which mainly depends on using a pre-trained network as a starting point to train a new task. This concept helps in making the process of training using a small number of samples more effective than training a network from scratch [80]. The purpose of transfer learning is to leverage the knowledge of a pre-trained model while performing a different task. Figure 3.38 from [81] shows the difference between the traditional machine learning technique and the transfer learning method. As in the transfer method, the learning process is faster, more precise, and needs fewer training samples and learning new task based on a pre-trained task. While on the other hand, in the traditional machine learning systems, each task is carried out separately, and no knowledge is transferred between tasks [82].

In our experiment, we used two well-known CNN architectures GoogleNet [83] and ResNet [84] to automatically extract the features from the handwritten documents. We then used the extracted features from each network individually and fed them to an SVM classifier to detect gender and age. Figure 3.42 shows the general structure of the automatic feature extraction.

The pre-processing step in this experiment is mainly the resizing of the images to the appropriate size related to each pre-trained network. For ResNet, we resized the images to (223x223x3), and for GoogleNet, we resized the images to (227x227x3). We applied the concept of transfer learning using two CNN architectures, as mentioned earlier as fixed feature extractors from the handwritten

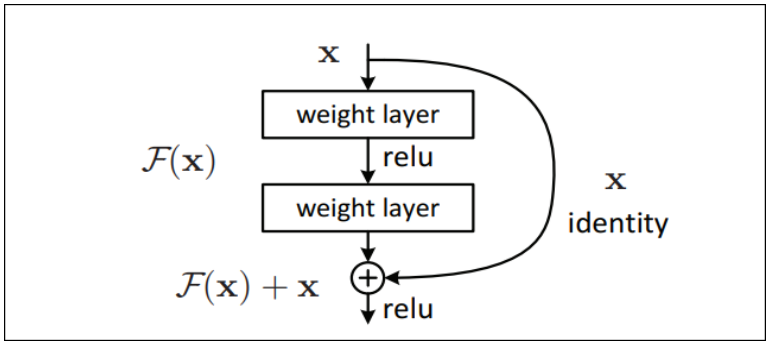


Figure 3.39: Residual learning: a building block

images. These CNN architectures have been trained over a million images from ImageNet, and they can classify images into 1000 categories. Training a network deeper makes the training processes very difficult. As the gradients become small, the weights become constant, leading to a halt in the network learning process, which affects the network's ability to do well. The residual network is a powerful way to solve the vanishing gradient problem. ResNet has residual blocks (skip connections) to jump over layers. Figure 3.39 taken from [84] shows the building block of ResNet, which speeds up learning by using a fewer number of layers to propagate through. The output of the skip connection is added to the stacked-layer to form $F(X) + X$. Figure 3.40 shows the general architecture of ResNet network. On the other hand, GoogleNet is an inception network. Therefore, it helps to reduce the computation cost by using a 1x1 convolution layer (bottleneck), which is used as a non linear dimension reduction module [83]. Figure 3.41 shows the building blocks of GoogleNet.

In their first convolutional layer, most convolutional neural networks learn to spot features like colour and edges. The network learns to detect more complex features as it progresses through the convolutional layers. The features of later layers are built up by merging the features of earlier ones. We applied the Stand-Alone Extractor approach, in which pre-trained layers are used to extract image features only once. The retrieved features would then be combined to generate a new dataset that does not require image processing.

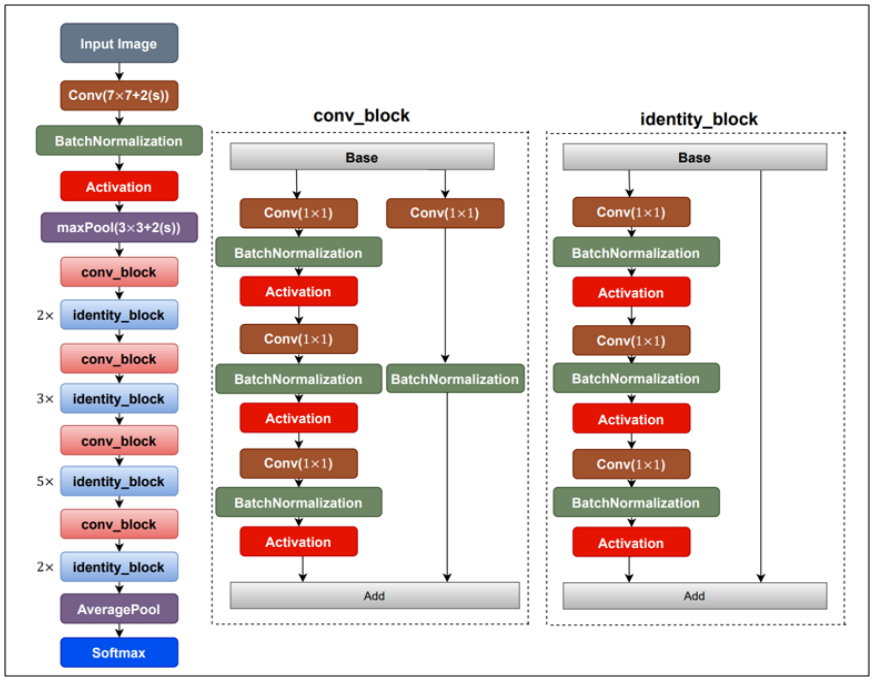


Figure 3.40: ResNet architecture

In our proposed work, The architecture of the pre-trained model was employed to extract features from our input dataset automatically. Only the Convolutional and Pooling layers were imported, but the ”upper portion” of the model was left out (the fully-Connected layer), which can be seen in Figure 3.42. Deeper layers of each CNN were used to extract features, which were then sent to an SVM classifier. We used the output of ”Pool5” layer and the ”inception_4e-1x1” layer from both ResNet and GoogleNet architectures, respectively.

3.4 Classification

Two classifiers were used to conduct the experiments of this research, one of which is Support Vector Machines (SVM) and the other one is Neural Network (NN). The following sections briefly describe the two classifiers.

Support vector machines [71] classify the input samples based on supervised learning methods. The strength of SVMs is centred on its ability to minimize the classification error and maximize the margins between separable classes. Only small numbers of the training data that can determine

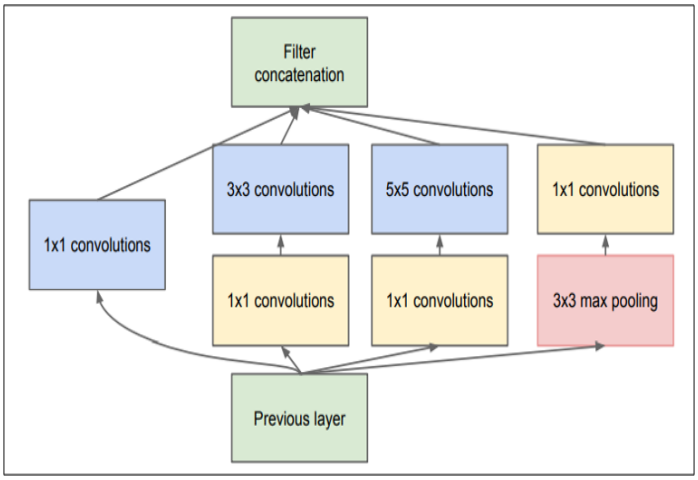


Figure 3.41: Building blocks of GoogleNet

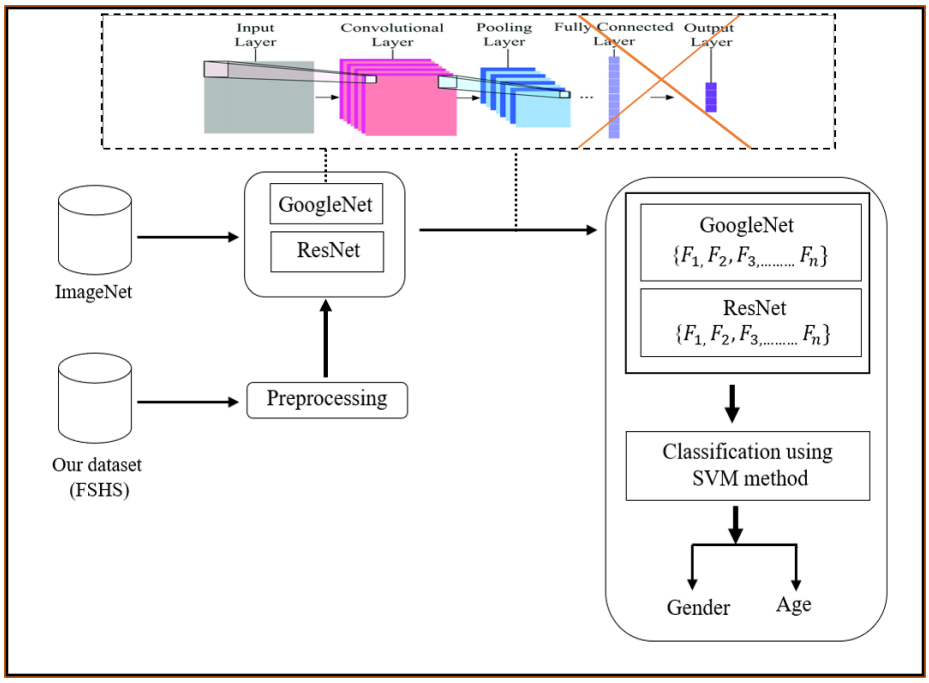


Figure 3.42: Structure of transfer learning method

this margin should be taken into account. These training data are called support vectors. SVM utilizes a non linear mapping kernel to map the non linearly separable vectors into a high dimensional feature space S where it can find the optimal hyper-plane to separate classes.

In our experiments, SVM was trained by a set of training samples labelled as 0 and 1 . As given a set of labeled training patterns $(v_1, y_1), (v_2, y_2) \ldots\ldots (v_n, y_n)$, where $y_i \in \{-1, 1\}$, $v_i \in V$, v is

p-dimensional training vector and V is the feature vector space, SVMs tend to solve the following optimization expression:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i$$
$$\text{subject to} \quad y_i(\mathbf{w}^T\phi(v_i) + b) \geq 1 - \xi_i, \qquad \xi \geq 0 \tag{12}$$

The function ϕ maps the training vectors vi to a higher dimensional space, and C is a penalty strength of the error term. ξ are slack variables which help the model to avoid overfitting on the training data. The idea of constructing support vector machines comes from considering the general form of the dot-products:

$$\phi(u).\phi(v) = K(u, v) \tag{13}$$

where $K(u, v)$ is called the kernel function. In our experiments, we used Gaussian kernel function:

$$K(v_i, v_j) = exp(-\frac{\|v_i - v_j\|^2}{2\sigma^2}) \tag{14}$$

where v_i , v_j are the support vectors and testing data point respectively and $\sigma > 0$ is the kernel parameter. These parameters were empirically chosen by using the validation set, which helps to find the optimal hyper-parameters. In this experiment, these parameters were optimally chosen using a Bayesian optimization search on the validation set.

Artificial Neural Networks (ANNs) [85] or Neural networks (NNs) is inspired by the sophisticated functionality of the human brain, where billions of interconnected neurons process information in parallel.

An artificial neural network consists of an input layer, hidden layer and output layer. Figure 3.43 shows the basic architecture of a neural network. As given a set of labeled training patterns $(x_1, y_1), (x_2, y_2) \ldots\ldots (x_n, y_n)$, where $y_i \in \{-1, 1\}$, $x_i \in V$, v is p-dimensional training vector and V is the feature vector space. The input neurons $(x_1, x_2, \ldots\ldots x_n)$ are all connected to the hidden

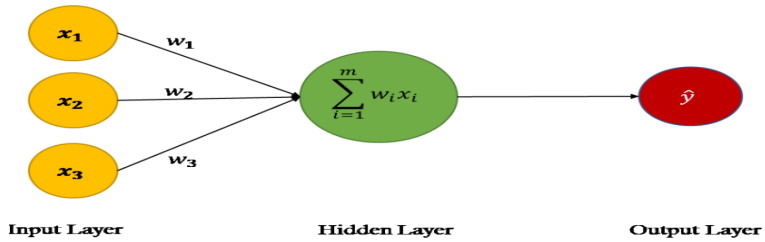


Figure 3.43: Basic Neural Network architecture

units. Each input neuron is connected to every hidden unit. The edges (connections) between them have a numerical number called weight (w_i), where i is the input vector.

NN networks tend to find the output by using an activation function. First, a sum of inner dot products of the input and their corresponding weights is calculated. Then it is fed to an activation function as follows:

$$\theta = f(b + \sum_{i=1}^{n} x_i w_i) \tag{15}$$

Many activation functions are extensively utilized, including relu, sigmoid, tanh and softmax. In our experiments, we used the following sigmoid activation function:

$$f(\theta) = \frac{1}{1 + e^{-}\theta} \tag{16}$$

The network training is based on finding the loss function, which is calculated with the actual expected output (y) and the predicted output ($\hat{y}$) using equation 17. The network tends to have the minimum error rate by tuning the weights and the biases. This process is called back propagation, which moves from the network's end to the beginning to find the most accurate output. In our experiments, the hyper-parameters were optimally chosen using a conjugate gradient back-propagation optimization search on the validation set.

$$Loss(y, \hat{y}) = \sum_{i=1}^{n} (y - \hat{y})^2 \tag{17}$$

Chapter 4

Experiments and Results

This section presents the experiments and the results obtained by the proposed systems. the experimental setup is described in section 4.0.1. The results are analyzed and discussed in section 4.0.2.

4.0.1 Experimental Setup

Gender Detection System Experimental Setup

The FSHS dataset (Section 3.1) consists of 2424 samples handwritten by 57% female writers and 43% male writers. 2000 samples, divided equally between female and male handwritten documents, were chosen to run the experiments and evaluate the proposed method where we followed the partitions that are mostly used in the handwritten analysis research which are 70% of the documents were used for training, 15% for testing and 15% for validation.

To compare the performance of our proposed method with other researcher works, we applied our system to the public Kaggle (ICDAR2013) dataset, consisting of 1900 images written by 475 writers in Arabic and English. We used both languages individually to train and test our system. Again, 70% of each category was used for training, 15% for testing and 15% for validation.

Figure 4.1 shows some handwritten samples of the ICDAR2013 dataset where Figure 4.1 (a)

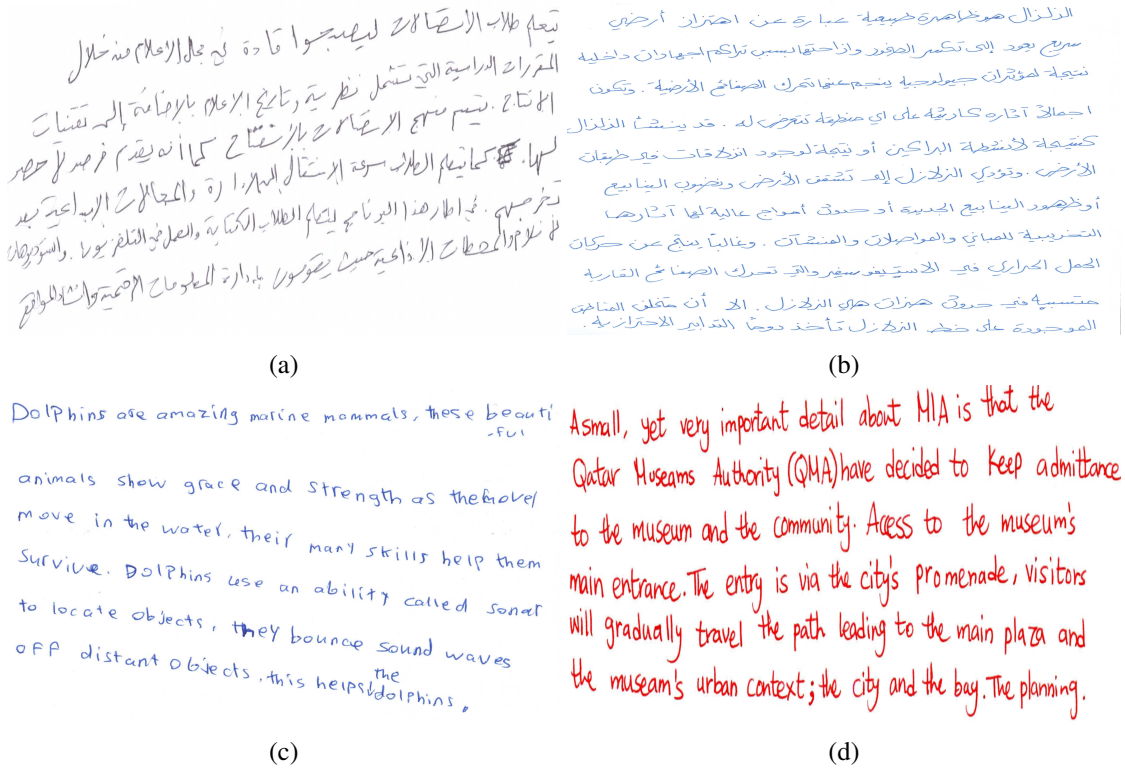


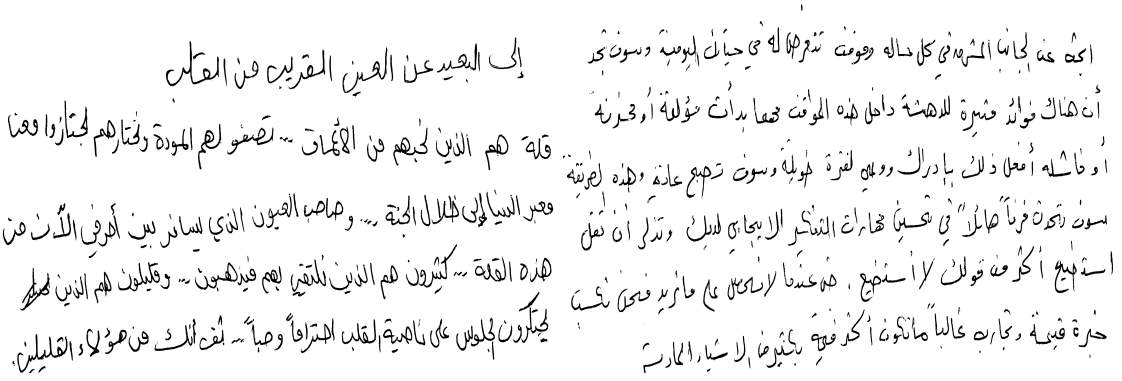
Figure 4.1: Handwritten samples from ICDAR2013 dataset

and (c) belong to a male writer, while Figure 4.1 (b) and (d) were written by a female.

Age Detection System Experimental Setup

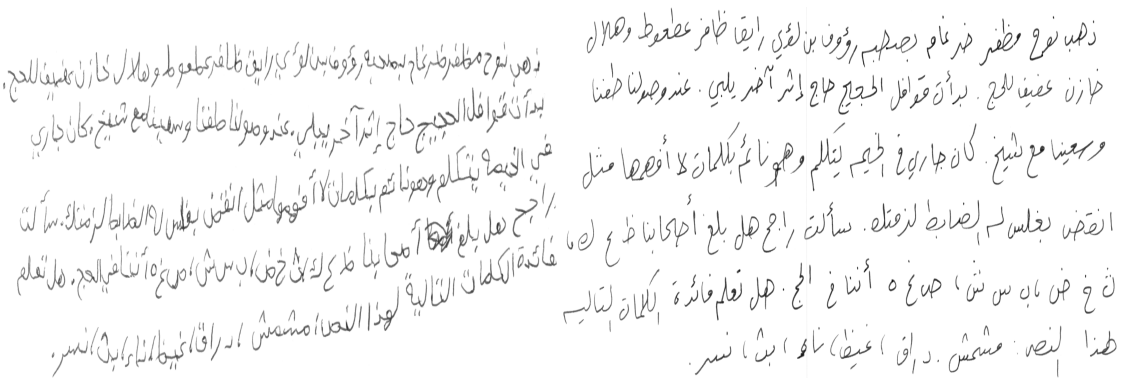
We used a subset of the FSHS dataset consisting of 2000 samples in our experiments. The images were divided into two main classes: youth adult class with ages ranging between 16 and 24 years old, and Mature adult class with ages ranging from 25 to 55 where each class has 1000 samples. Figure 4.2 (a) shows a document written by an adult youth writer with slant and text lines irregularity, while Figure 4.2 (b) shows mostly regular writing of a mature adult writer. To run the experiments and evaluate the proposed method, 70% of the documents were used for training, 15% for testing and 15% for validation.

To benchmark the performance of the proposed method with other age detection methods, we applied our system to the public Khatt [40] dataset. That dataset contains Arabic handwritten documents with ground truth of gender, age, and handedness. It has three age ranges that can be defined



(a) (b)

Figure 4.2: Samples of handwritten documents from FSHS dataset



(a) (b)

Figure 4.3: Samples of handwritten documents from Khatt dataset

as: ”under 15 years old”, ”16 to 25 years old”, and ”26 to 50 years old”. Each class contains 135 samples. Similar to the available works on the khatt dataset, we adapted two age ranges, ”16-25” and ”26 and above”. Again, 70% of the dataset was used for training, 15% for testing and 15% for validation. Figure 4.3 shows some handwritten samples of the Khatt dataset where Figure 4.3 (a) belongs to a youth adult, while (b) was written by a mature adult.

4.0.2 Evaluation

Any project must include an evaluation of a machine learning method. As a result, specific performance metrics are required to evaluate any system. Accuracy, recall, and precision are three popular approaches to assess a handwriting analysis system’s performance, whether using correctly categorised samples or wrongly classified data [86]. The following formulas can be used to compute

the values of various performance metrics:

Accuracy is the most common performance metric; it measures the ratio of correct predictions to the total number of input samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

Recall is the number of correct positives divided by the number of all samples that were classified correctly. it is used to evaluate the model's quantity performance. as it indicates that the algorithm returns the majority of the relevant outcomes.

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

Precision is the number of correct positive results divided by the number of positive results predicted by the classifier. It is typically used to calculate the cost of the FP findings because it indicates the quality of the model. A more precise algorithm will produce more relevant results than irrelevant ones.

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

where, TP (True Positive): Determined as a document being classified correctly as relating to a category. FN (False Negative): Determined as a document that is not marked as related to a category but should be. FP (False Positive): Determined as a document that is said to be related to the category incorrectly. TN (True Negative): Documents that should not be marked as being in a particular category and are not.

Gender Detection System Evaluation

Our experiment is a two-class classification problem where the outcomes are either a male writer or a female writer. We plot the predicted values (results) M against the ground truth (actual) values N to get the confusion matrix [87] $M \times N$, which helps to evaluate the performance of supervised

Actual class	Predicted class (female)	Predicted class (male)
Female	TP	FP
Male	FN	TP

Table 4.1: Gender detection confusion matrix

learning algorithms. Table 4.1 shows the confusion matrix of the gender detection system, which can be used to derive several performance metrics such as accuracy, precision, recall and AUC (Area Under Curve). Our experiments were divided into four categories: individual features, two-features combinations, three-features combinations, and all-features combination.

Table 4.2 shows the experimental results of the proposed method in terms of precision, recall and accuracy using SVM and NN, respectively. We evaluated the performance of each feature individually to understand the relevance of each one. It can be shown that all the features attain a high performance when applied to our dataset (FSHS). Moreover, among these features, it appears that the text line irregularity feature can detect gender of the writer with high precision and recall rates. Pen pressure has an accuracy of 76% which is higher than word-spacing and margins, although it can be calculated using a feature vector of only two features for each document, in contrast to the word-spacing with 32 features and the margins that have 400 features. Despite of the variance in each feature performance, the combination of all features resulted in a high accuracy of 94.7% using an SVM classifier with the Gaussian kernel and an accuracy rate equal 97.1% using ANN.

On the other hand, for the two features combinations, it can be noticed that the combination of the word spacing and the irregularity has the highest performance among the others for both SVM and NN. While the combination of margins, irregularity and pen pressure has relatively a higher accuracy rate than the other three features combinations.

Extracted features				SVM classifier					NN classifier				
Margins	Word Spacing	Irregularity	Pen Pressure	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Female	Male	Female	Male		Female	Male	Female	Male
✓	×	×	×	73.8	67.7	100	100	41.6	60.5	60.6	60.5	66.8	53.8
×	✓	×	×	74.7	68.4	100	100	44.2	56.5	76.2	31.5	58.6	51.0
×	×	✓	×	87.8	82.1	99.5	99.7	73.3	89.4	97.1	81.1	84.7	96.3
×	×	×	✓	76	84.2	69.2	69.5	84	62.4	95.5	13.8	61.9	67.7
✓	✓	×	×	73.3	67.2	100	100	41	62.1	61.7	62.7	66.8	57.3
✓	×	✓	×	91.8	89.7	94.9	96.2	86.5	95.7	93.6	98.1	98.2	93.4
✓	×	×	✓	73.4	73.4	73.6	81	64.2	62.7	55	68.9	59.2	65.1
✓	✓	✓	×	92.3	88.7	98	98.5	84.8	96	97.3	94.4	95.8	96.4
✓	×	✓	✓	92.8	89.2	98.4	98.8	85.5	96.1	95.7	96.5	97.3	94.6
✓	✓	×	✓	76.2	70.4	94.6	97.7	49.8	61.3	67	53.8	66.1	54.8
×	✓	✓	×	93.6	90.4	98.4	98.8	87.2	96.4	97.8	94.6	95.7	97.2
×	✓	×	✓	74.1	68	100	100	42.3	54.7	86.8	22	53.1	62.1
×	✓	✓	✓	92.8	89.4	98	98.5	85.8	95.9	95.6	96.4	97.2	94.4
×	×	✓	✓	90.9	87.4	96.3	97.4	82.9	95.2	93.6	96.6	96.1	94.5
✓	✓	✓	✓	94.7	98.4	92.1	89.7	98.8	97.1	96.8	97.3	96.8	97.3

Table 4.2: Results of applying the proposed Gender Detection method to FSHS dataset

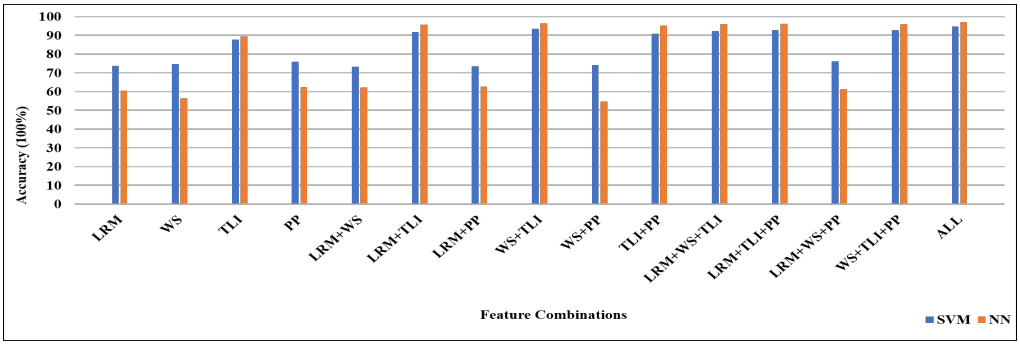


Figure 4.4: Comparison of feature importance between SVM and NN

Figure 4.4 presents the obtained results in terms of accuracy rates. We can notice that there is stability in the importance of features between the SVM and NN classifiers. The text line irregularity (TLI) feature has the best performance among the separate features. While comparing to the other two-features combinations, we can see that three of them: (LRM + TLI), (WS + TLI) and (TLI + PP) have good classification rates. While looking at the three-features combinations, we can observe that (LRM + WS + PP) has the lowest performance of the four combinations.

The performance of the proposed gender detection system with its four feature categories was also evaluated using the Receiver Operating Characteristic curve (ROC) [88] and AUC. ROC curve is a plot that represents the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) at various threshold levels. While the AUC measures a model's entire performance, the greater the AUC value, the better the model.

The ROC curve and AUC for the four gender-relevant features are shown in Figure 4.5. As (c) belongs to the text-line irregularity (TLI) feature, the findings are compatible with precision, recall, and accuracy metrics, with the maximum AUC of 0.94 and 0.92 for both SVM and NN, respectively.

Moreover, Figure 4.6 shows the ROC curve and AUC for the two-features category. Again, the three combinations, as shown in (b), (d), and (f), had the best performance among the others, with AUC values ranging from 0.98 to 1 for both SVM and NN methods.

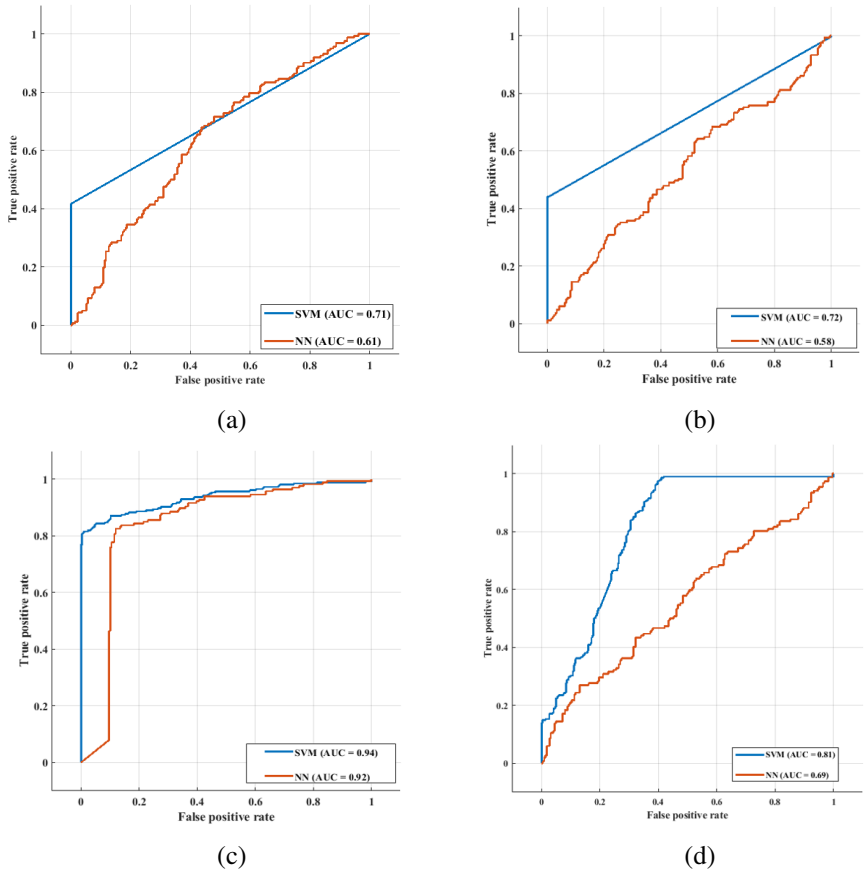


Figure 4.5: ROC comparison between SVM and NN results using individual features on (a) Margins, (b) Word spacing, (c) Irregularity, and (d) Pen pressure, all applied on FSHS dataset.

Furthermore, among the other three-feature combinations, (LRM + WS + PP) has the lowest performance, with AUC values of 0.76 and 0.64 for SVM and NN, respectively. Figure 4.7 illustrates this conclusion.

The classification rate increases when all the features are integrated. For all feature combinations using SVM method, Figure 4.8 (a) shows the ROC curve and AUC of 0.99, whereas Figure 4.8 (b) shows the optimum hyperparameter point reached on iteration 26 with the lowest classification error of 0.02. The full feature combination hit an accuracy rate of 94.7% using the SVM method.

Figure 4.9 represents the results of using the full features NN model. Figure 4.9 (a) depicts the ROC curve and AUC value of 1, whereas Figure 4.9 (b) depicts the training NN characteristics

in terms of minimum errors in training, validation, and testing, which tends to be stable after the 39th epoch generation, indicating that the network convergence speed is fast and the network performance is good.

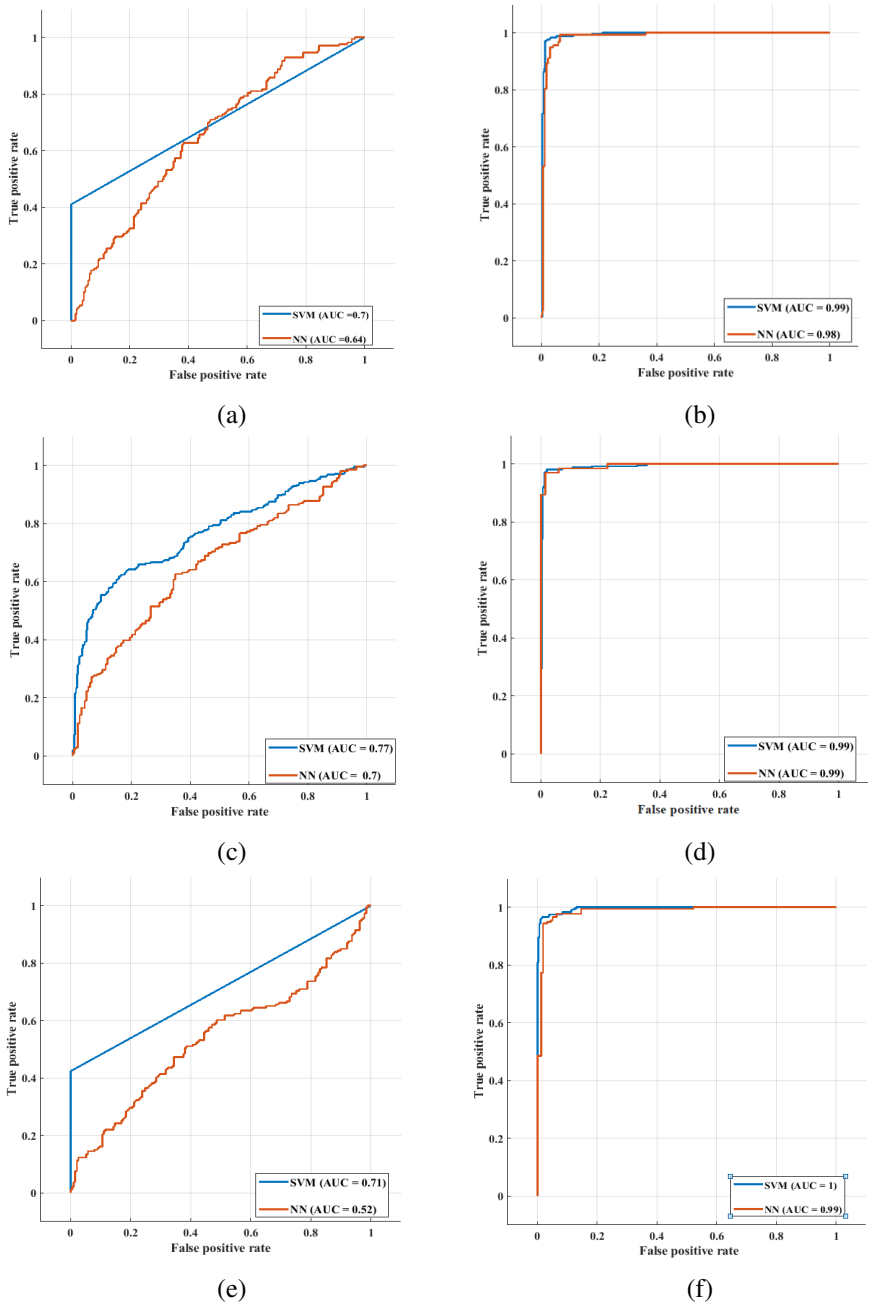


Figure 4.6: ROC comparison between SVM and NN results using two features combination on (a) LRM + WS, (b) LRM + TLI, (c) LRM + PP and (d) WS + TLI, (e) WS + PP, and (f) TLI + PP, all applied on FSHS dataset.

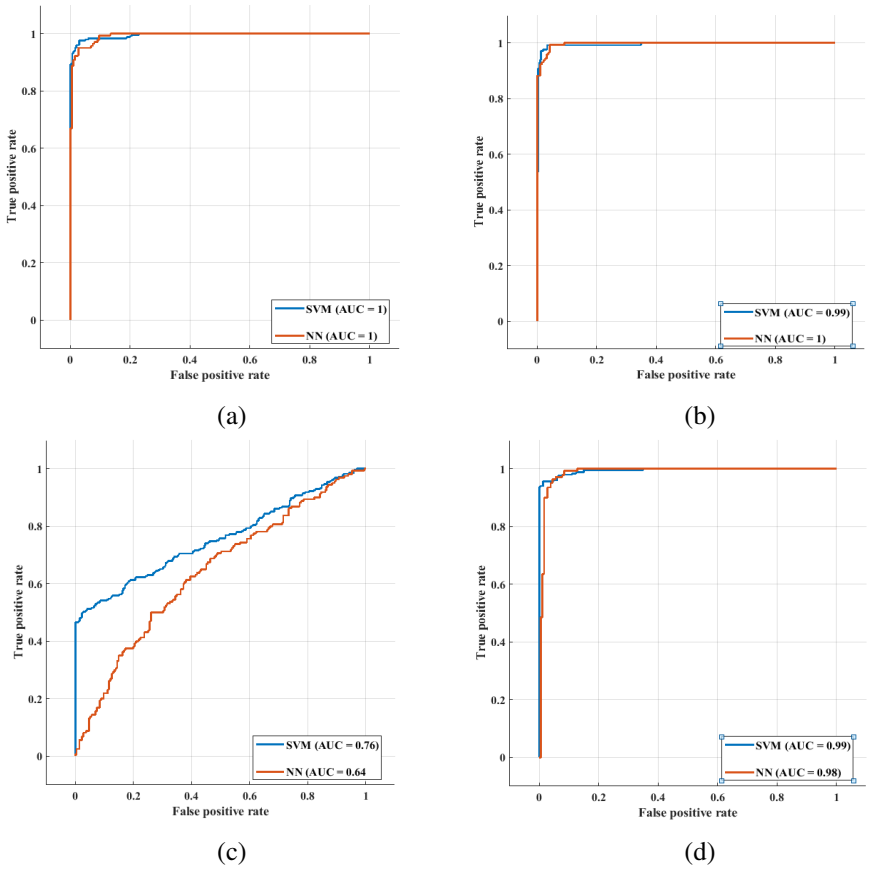


Figure 4.7: ROC comparison between SVM and NN results using three features combinations on (a) LRM + WS + TLI, (b) LRM + TLI + PP, (c) LRM + WS + PP, and (d) WS + TLI + PP, all applied on FSHS dataset.

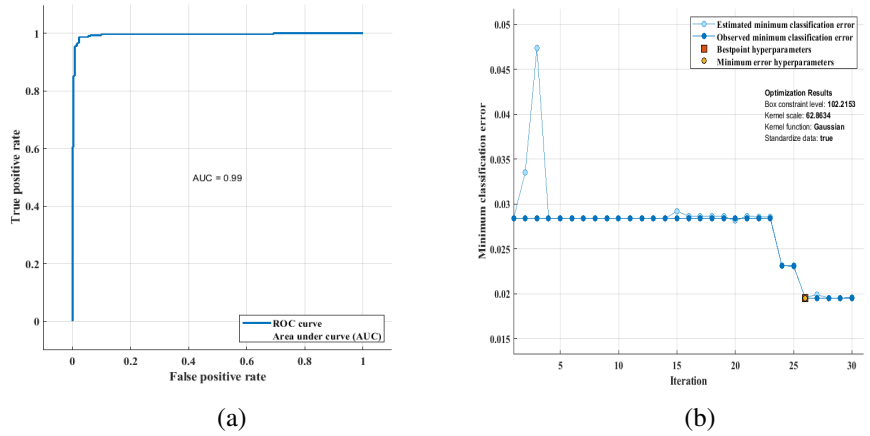
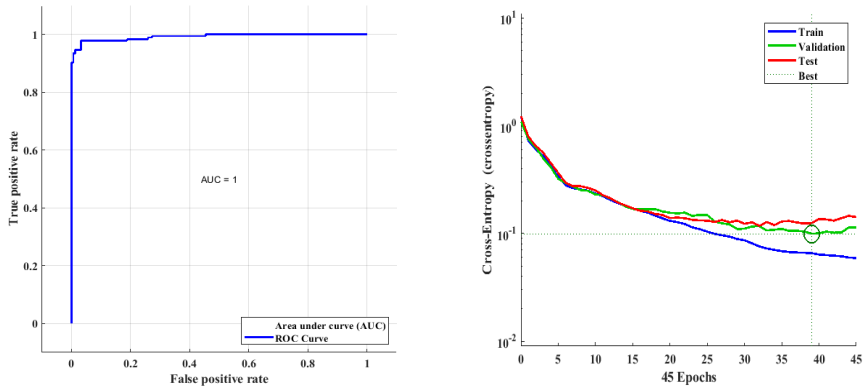


Figure 4.8: Performance of the SVM model using the full features combination in terms of (a) ROC curve, and (b) The minimum classification error.



(a) (b)

Figure 4.9: Performance of the NN model using the full features combination in terms of (a) ROC curve, and (b) Loss function.

To compare our proposed method to other existing methods, we used the public dataset ICDAR2013 to test our gender detection system. The results of implementing the proposed manual feature extraction method utilising both SVM and NN methods on the Arabic and English subsets of the ICDAR2013 dataset are presented in Tables 4.3 and 4.4, respectively.

Individual features and all feature combination were examined on both subsets. With findings similar to those obtained from the FSHS dataset, the text lines irregularity (TLI) has the highest accuracy using the Arabic subset, with 91.7% and 88.7% using SVM and NN, respectively, followed by pen pressure with 65.4% and 61.7% using SVM and NN. Word spacing and margins were ranked third and fourth, respectively.

Using the SVM algorithm on the Arabic subset, the proposed strategy outperformed the available methods by 96.2% of accuracy rate. Table 4.5 shows how our proposed methods compare to the current state-of-the-art.

Extracted Features	SVM Classifier					NN Classifier				
	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
		Female	Male	Female	Male		Female	Male	Female	Male
Text line irregularity (TLI)	91.7	95.1	88.9	87.9	95.5	88.7	94.4	82	86.1	92.6
Pen Pressure (PP)	65.4	60.6	61.3	64.2	57.6	61.7	66.7	56.7	60.3	63.3
Word Spacing (WS)	63.2	62.7	60.8	56.1	67.2	60.2	47.6	71.4	60	60.2
Left and Right Margins (LRM)	61.7	60.2	68	75.8	50.7	55.7	60.6	50	58.1	52.5
ALL Features (TLI +PP+WS +LRM)	96.2	98.5	94	94.2	98.4	92.5	91.2	93.8	93.9	91

Table 4.3: Results of applying the proposed gender detection system on Arabic subset of ICDAR2013 dataset.

Extracted Features	SVM Classifier					NN Classifier				
	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
		Female	Male	Female	Male		Female	Male	Female	Male
Text line irregularity (TLI)	84.2	83.6	84.9	85.8	82.6	74.6	78.7	69.5	76.6	71.9
Pen Pressure (PP)	57.1	58.3	56.2	52.2	62.1	53.7	55.7	51.6	55.7	51.6
Word Spacing (WS)	62.2	61.1	63.7	70.8	53.2	58.2	56.5	60	60	56.5
Left and Right Margins (LRM)	61.7	62.3	60.9	63,2	60	56.7	76.1	37.3	54.8	61
ALL Features (TLI +PP+WS +LRM)	88.7	87.1	90.5	91	86.4	83.6	87.5	79	82.9	84.5

Table 4.4: Results of applying the proposed gender detection system on English subset of ICDAR2013 dataset.

The results of applying the proposed gender detection method to the English subset of the ICDAR2013 dataset are shown in table 4.4. It is worth noting that, once again, the text line irregularity feature outperforms the others, with accuracy rates of 84.2 and 74.6, respectively, using the SVM and NN approaches. Word spacing, on the other hand, has the second-highest accuracy rates, with 62.2 and 58.2 using SVM and NN approaches, respectively. Despite this, the pen pressure has the lowest accuracy among the other accuracy rates, with 57.1 and 53.7 for SVM and NN, respectively.

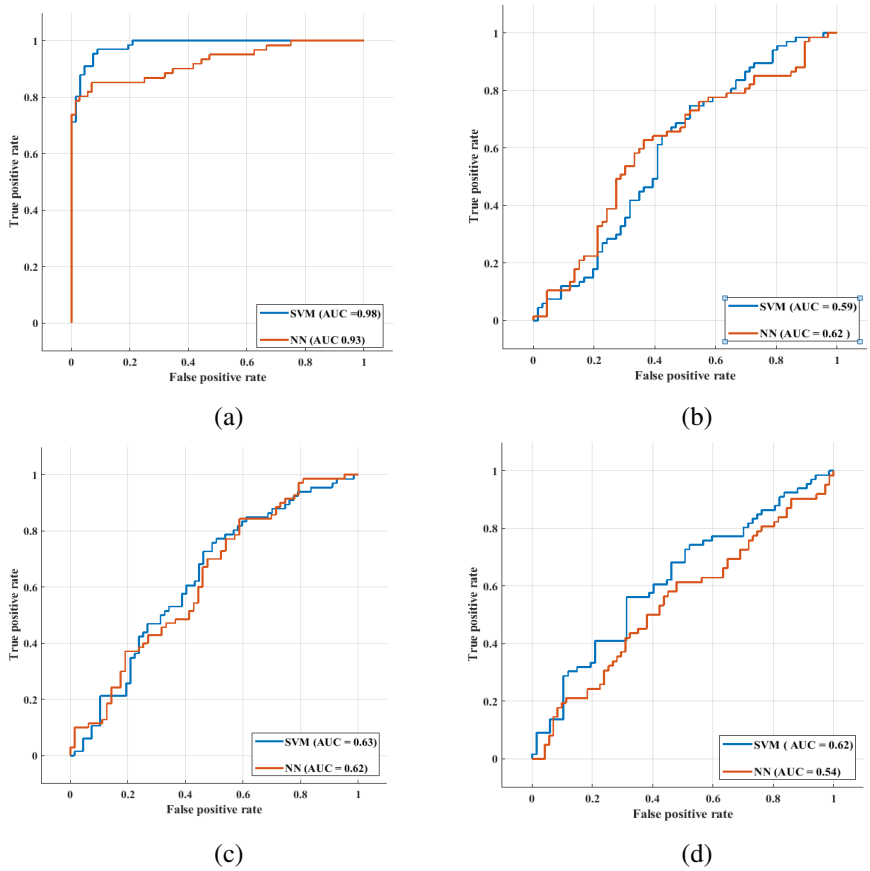


Figure 4.10: ROC comparison between SVM and NN results using individual features models: (a) TLI, (b) PP, (c) WS, and (d) LRM applied on Arabic subset of ICDAR2013 dataset.

AUC was also used to test the proposed gender detection method using the ICDAR2013 dataset. The ROC curves and AUC values obtained by applying the proposed models: (a) TLI, (b) PP, (c) WS, and (d) LRM on the Arabic subset are shown in Figure 4.10. The TLI clearly outperforms the other features with AUC values of 0.98 and 0.93 using SVM and NN, respectively.

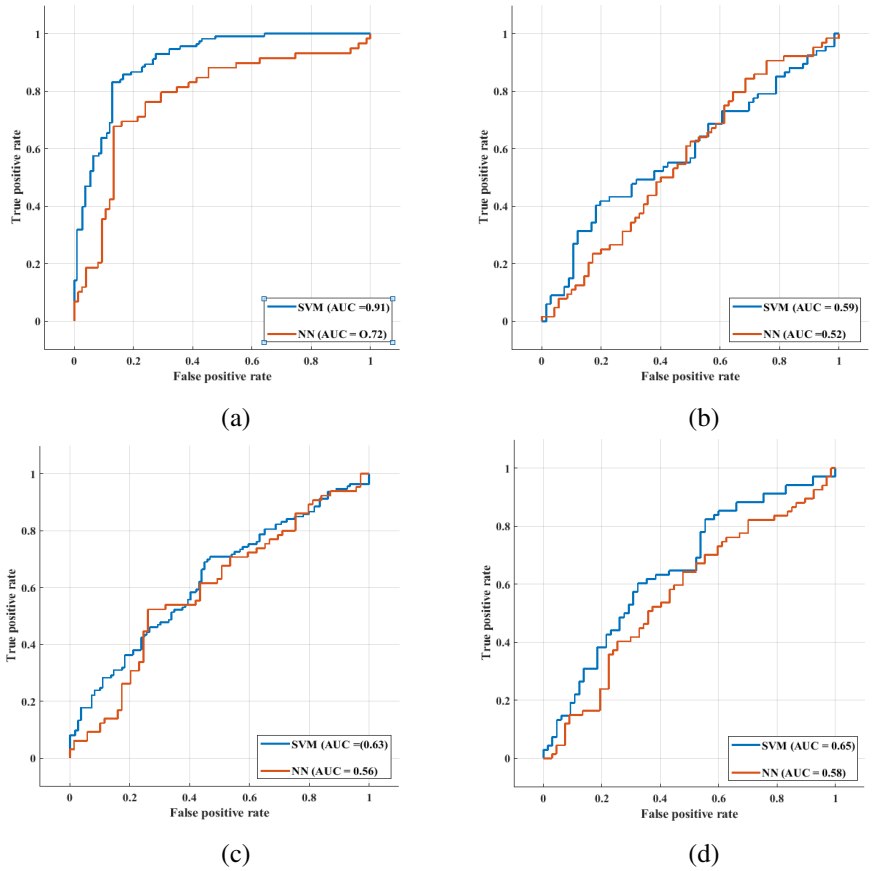
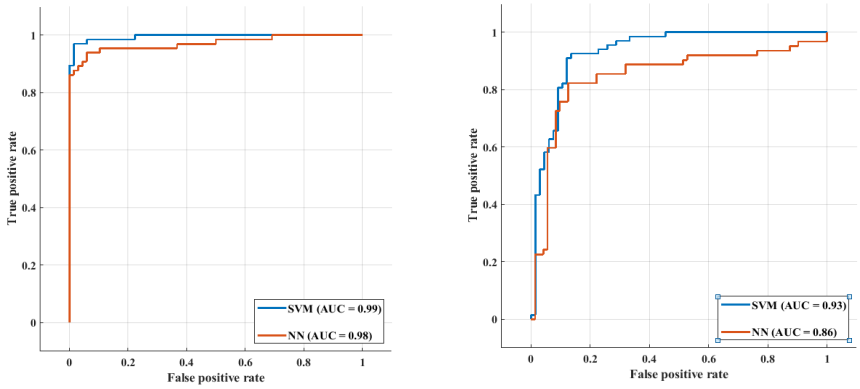


Figure 4.11: ROC comparison between SVM and NN results using individual feature models: (a) TLI, (b) PP, (c) WS, and (d) LRM applied on English subset of ICDAR2013 dataset.

The AUC values presented in Figure 4.11 were obtained by applying the proposed gender detection to the English subset of the ICDAR2013 dataset. It can be shown that utilising SVM and NN, the TLI has the best performance with AUC values of 0.91 and 0.72, respectively.

Furthermore, Figure 4.12 displays the ROC and AUC values of the proposed full features applied to Figure 4.12 (a) the Arabic subset using both SVM and NN, with AUC values ranging from 0.98 to 0.99, and Figure 4.12 (b) which demonstrates the performance of the full features combination model with AUC values of 0.93 and 0.86 using SVM and NN, respectively, applied on the English subset.



(a) (b)

Figure 4.12: AUC values of all features combination model applied on (a) Arabic subset and (b) English subset.

Method	Langauge		Classifier	Accuracy (100%)
	Training	Testing		
Siddiqi et al.(2012)	Arabic	Arabic	SVM	62.3
	English	English		77.1
Youssef et al.(2013)	Arabic	Arabic	SVM	68.5
	English	English		68.5
Hassaine et al. (2013)	Arabic	Arabic	SVM	68.6
	English	English		85.7
Akbari et sl.(2017)	Arabic	Arabic	SVM	77.7
	English	English		75.5
Gattal et al. (2018)	Arabic	Arabic	SVM	76.17
	English	English		77.98
Bi et al. (2019)	Arabic And English	Arabic And English	KMI	66.3
Xue et al. (2020)	Arabic And English	Arabic And English	ATP-DenseNet-169	71.8
Rabaev et al. (2021)	Arabic	Arabic	Deep Learning	50
	English	English		67
Fahimeh and Alireza (2021)	Arabic	English	SVM	73
	English	Arabic		70
Maken and Gupta (2021)	English	English	SVM	61.5
	English	English	Logestic Regresion	63.3
	English	English	KNN	56.5
	English	English	Majority Voting	65.71
proposed all features combination	**Arabic**	**Arabic**	**SVM**	**96.2**
			NN	**92.5**
	English	**English**	**SVM**	**88.7**
			NN	**83**
Proposed text line irregularity (TLI)	**Arabic**	**Arabic**	**SVM**	**91.7**
			NN	**88.7**

Table 4.5: Comparison between the results of our proposed gender detection method with other research methods applied on ICDAR2013 dataset.

The proposed model outperforms the state of the art when using Arabic and English datasets. Table 4.5 compares our proposed gender detection method with the existing state of the art. Two proposed models outperform the existing works. The first is the text line irregularity (TLI) feature, with an accuracy rate of 91.7% when using SVM and 88.7% when using NN, both applied on Arabic subset. The second proposed model contains all features that surpass available research using SVM and NN on the Arabic subset, with accuracy rates of 96.2% and 92.5%, respectively, and an accuracy rate of 88.75% using SVM on the English subset.

In addition to employing classic image processing techniques to extract the features, we employed the transfer learning technique to automatically extract the features from the handwritten documents using multiple CNN architectures.

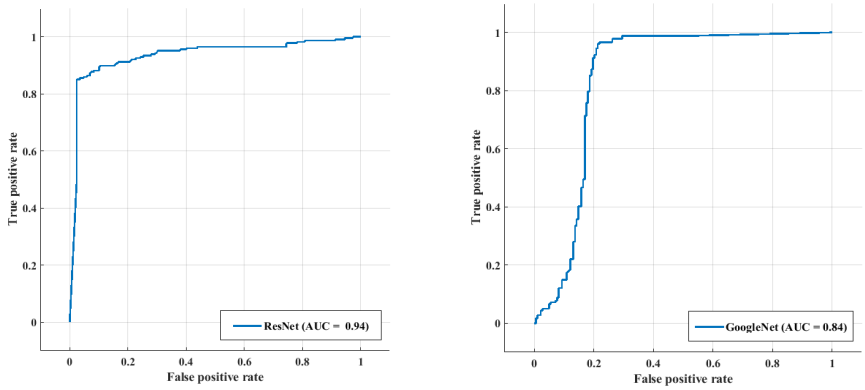
Table 4.6 shows the accuracy, precision, and recall values achieved from applying transfer learning to the FSHS dataset using ResNet and GoogleNet architectures. ResNet attained an accuracy of 84.9%, which is higher than the accuracy reached by GoogleNet, which is 82.2%. Figure 4.13 also shows the ROC and AUC values for the two architectures. For example, the ResNet in (a) has a greater AUC value of 93.5%, whereas the googleNet in (b) has an AUC value of 84.35%.

CNN	Accuracy (100%)	Precision (100%)		Recall (100%)	
		Female	Male	Female	Male
ResNet	84.9	77.7	92.1	90.8	80.5
GoogleNet	82.2	65.6	98.9	98.4	74.1

Table 4.6: Results of applying transfer learning method to the FSHS dataset

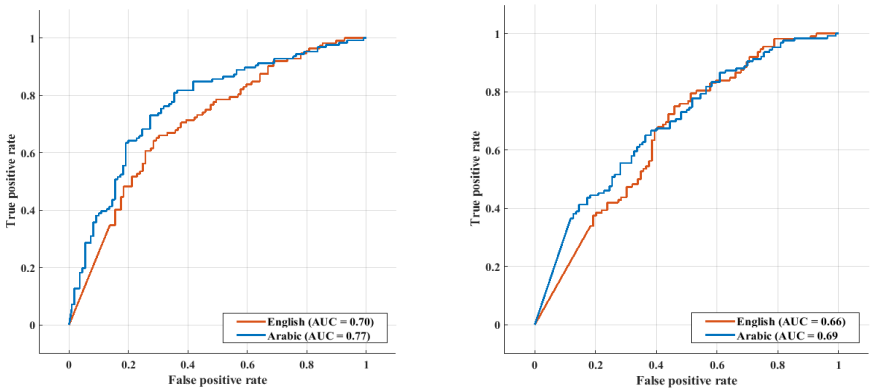
CNN	Language	Accuracy (100%)	Precision (100%)		Recall (100%)	
			Female	Male	Female	Male
ResNet	English	66.06	67.9	64.2	66.1	66
	Arabic	72	75.4	68.2	73.1	70.8
GoogleNet	English	63.3	67.9	58.7	62.8	64
	Arabic	64	66.7	60.9	66.1	61.5

Table 4.7: Results of applying transfer learning method to the ICDAR2013 dataset



(a) (b)

Figure 4.13: ROC curves and AUC values of (a) ResNet and (b) GoogleNet applied to the FSHS dataset



(a) (b)

Figure 4.14: ROC curves and AUC values of (a) ResNet and (b) GoogleNet applied to the ICDAR2013 dataset

Moreover, the experimental results of implementing the proposed automatic feature extraction methods on the ICDAR2013 (Kaggle) dataset are also shown in Table 4.7. When applied to English and Arabic subsets, the ResNet network had a better accuracy rate of 66.06% and 72%, respectively. The results of ResNet in terms of accuracy outperform the results obtained by Xue et al. [61] and Rabaev et al. [62], both of which used deep learning in their methodologies. The two CNN architectures were also evaluated using ROC and AUC on the ICDAR2013 dataset. The ROC curves of (a) ResNet and (b) GoogleNet are shown in Figure 4.14. When applied to the Arabic and English subsets, the ResNet has an AUC of 0.77 and 0.70, respectively. While the GoogleNet has an AUC values of 0.69 and 0.66 when applied on Arabic and English subsets, respectively.

What can be inferred from the obtained results is that the traditional methods of extracting features produce higher classification results than the automatic method using transfer learning. As the extracted features in the second method are not directly related to the characteristics of the writer. Moreover, it can be noticed that the transfer learning method gives a higher classification rate when applied to a larger dataset, where we got a higher accuracy rate when we applied it to our dataset (FSHS) compared to the ICDAR2013 dataset.

Age Detection System Evaluation

Our study is a two-class classification problem, with the outcomes being either a young adult writer or a mature adult writer. The accuracy, precision, and recall metrics were used to evaluate the proposed method which can be calculated using Equations 18, 19 and 20 mentioned in Section 4.0.2.

To produce the confusion matrix $M \times N$, we plot the predicted values (results) M against the ground truth (actual) values N. Table 4.8 depicts the age detection system's confusion matrix, which was utilised to calculate the performance metrics.

Actual class	Predicted class	
	Youth (adult)	Mature (adult)
Youth (adult)	TP	FP
Mature (adult)	FN	TN

Table 4.8: Age detection confusion matrix

The experimental results of the proposed age detection approach in terms of precision, recall, and accuracy utilising SVM and NN methods are shown in Table 4.9 and Table 4.10. Individual features, three-feature combinations, and all-feature combinations were the three main categories in which we split our experiments. All the feature combinations results are available in A. The comparison of different feature combinations between SVM and NN is shown in Figure 4.15. We can see that the relevance of features is consistent between the SVM and NN classifiers.

Extracted features				SVM classifier					NN classifier				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)		Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)
✓	×	×	×	56.3	56.1	56,6	58	54.7	56.8	49.2	63.5	54.2	58.8
×	✓	×	×	53.8	54.4	53.4	49.2	58.5	52.7	98.7	6.0	51.6	81.6
×	×	✓	×	66.2	64.7	68.2	71.8	66.7	58.5	44.9	73.1	64.2	55.2
×	×	×	✓	55.7	63.4	53.5	27.5	83.9	55.7	34.6	77.6	61.6	53.3

Table 4.9: Results of individual features models applied to FSHS dataset.

Extracted features				SVM classifier					NN classifier				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)		Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)
✓	✓	✓	×	70.7	69.1	72.6	75.5	65.8	62.3	62.2	62.5	66.7	57.8
✓	×	✓	✓	67.3	66.7	68.1	70.2	64.4	59.2	66.2	52.4	57.5	61.5
×	✓	✓	✓	69.7	68.8	70.6	72	67.3	60	64.0	56.0	59.3	61
✓	✓	✓	✓	71	68.9	73.7	76.7	65.3	63.5	64.6	62.3	66.0	60.9

Table 4.10: Results of The best features combination models applied to FSHS dataset

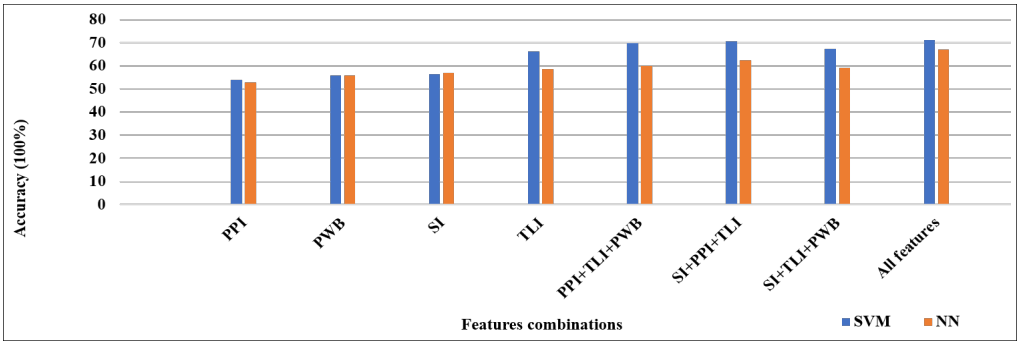


Figure 4.15: Comparison of feature importance between SVM and NN

We firstly evaluated the performance of each feature individually to understand the relevance of each one. Using the SVM and NN methods, we can find that the TLI has a greater accuracy rate than the other features, with 66.2% and 58.5%, respectively. Additionally, The TLI model performs well in the mature adult writer category because only 25.3% of the mature writers handwritten samples were incorrectly categorised as youth adult writers, as opposed to 42.7% of youth writers samples were incorrectly classified as mature adult writers. After the TLI, the SI feature ranked second with an accuracy rate of 56.3% using SVM and 56.8% using NN. Furthermore, 45% of youth samples were miss classified as mature adults using SI model, and 42% of the mature samples were incorrectly classified as youth-adults. PWB and PPI features are ranked third and fourth, respectively. 80% of the mature adult samples were classified correctly using PWB model while it was 51.3% of the samples using PPI. These findings indicate that the TLI is a useful feature in handwritten analysis applications, as it performs well in gender and age detection. In addition, PWB has the highest performance when applied to the mature adult category which makes it a powerful feature for old age groups application. Last but not least, the majority of samples that are correctly categorised fall mostly into the mature adult group, showing that older writers have more consistent and reliable writing styles than youth adult writers.

The best results from combining the individual features are presented in Tables 4.10. In the three feature combination studies, the SI + PPI + TLI combination yielded the highest accuracy rates of 70.7% and 62.3% using SVM and NN, respectively. 24.5% of the samples from youth writers and 24.8% from mature writers were miss classified by this approach. The PPI + TLI + PWB model

came in second with 69.7% and 60% accuracy rates using SVM and NN, respectively. In addition, 72% of the youth writers' samples and 67% of the mature writers' samples were correctly classified using this combination. Using SVM and NN, SI + TLI + PWB was ranked at the end of the three feature combinations with accuracy rates of 67.3% and 59.2%, respectively. On the other hand, 29.8% of youth writers' samples and 35.6% of mature writers' samples were miss classified.

Combining all features improves the proposed system's performance, as shown in Table 4.10, where the accuracy rates of all features models utilising SVM and NN are 71% and 63.3%, respectively, which puts it ahead of all other feature combinations. 23.3% and 34.7% were the percentage of miss classified samples in both categories youth and mature writers samples, respectively.

According to our manual error analysis of all the models, the majority of the miss-classified samples for all of the proposed traditional feature extraction's age classification models applied to the FSHS dataset occur in the age range of 21 to 28 years old, which is considered the transitional period between the two main proposed classes (youth-adult and mature adult). The remaining miss classified samples were distributed among the other age groups as a result of the significant similarity in the handwriting extracted features from different writers. Despite having an error rate that ranges from 29% to 46.3%, The proposed models can categorise the handwritten samples in these categories with an accuracy of between 55% and 62%. Gathering more data samples for the instances when mistakes occurred is one of the suggested future improvement for the proposed system.

The proposed age detection system's performance with its four feature categories was further assessed using the Receiver Operating Characteristic curve (ROC) and the Area Under Curve (AUC). Figure 4.16 presents the ROC curves and AUC values of the individual age-relevant feature models. The AUC values are consistent with the accuracy rate results, as the TLI in (c) has a higher AUC value than the other features, which is 0.68 and 0.61 using SVM and NN, respectively. The SI feature in (a) is next, with AUC values of 0.56 and 0.59 using SVM and NN techniques. On the other hand, PPI in (b) and PWB in (d) have similar AUC values, putting them in the last rank.

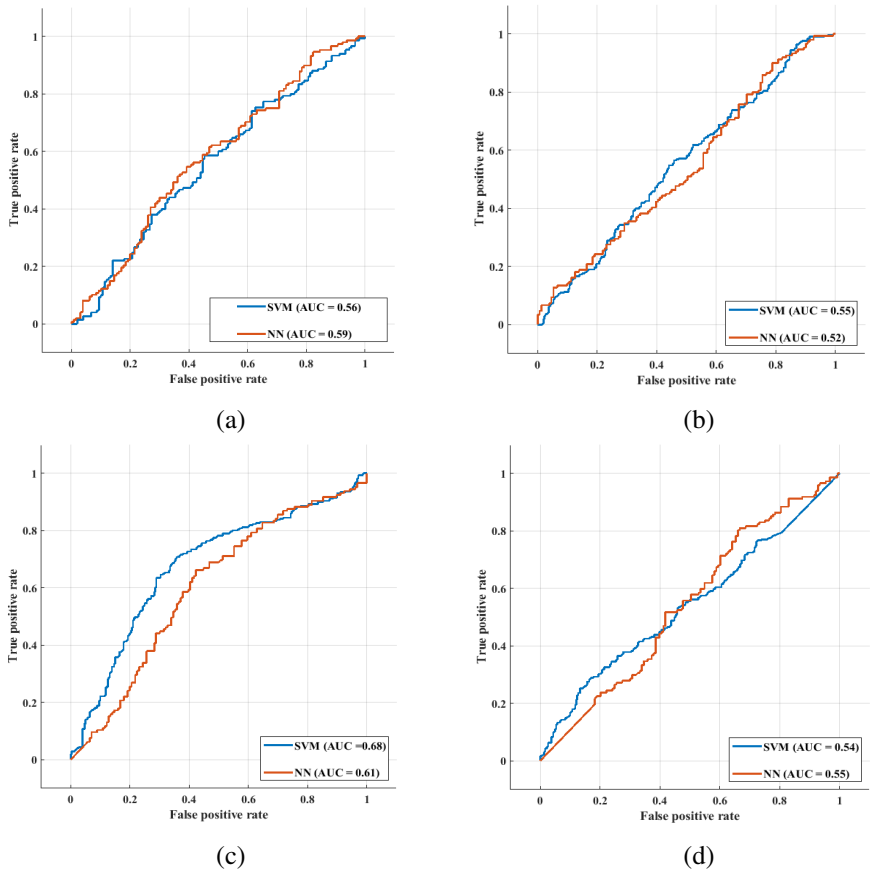


Figure 4.16: ROC curves and AUC values using individual feature models: (a) SI, (b) PPI, (c) TLI, and (d) PWB, all applied to FSHS dataset

Moreover, SI + PPI + TLI, as shown in Figure 4.17, has the best AUC values, which correlates to the findings of the three feature combinations provided in Table 4.9, as the AUC values for the SVM and NN methods are 0.77 and 0.67, respectively. Additionally, With AUC values of 0.72 and 0.64 using SVM and NN, the PPI + TLI + PWB. On the other hand, SI + TLI + PWB model has an AUC values of 0.74 and 0.62 using SVM and NN, respectively.

Figure 4.18 (a) depicts the ROC-AUC of all features combination model with an AUC of 0.78 using the SVM classifier, whereas Figure 4.18 (b) depicts the minimum classification error of 0.26 on iteration 30. Figure 4.19 also illustrates the ROC curve and loss function for the all-features combination using the NN method. As seen, the best validation point occurred at epoch 46.

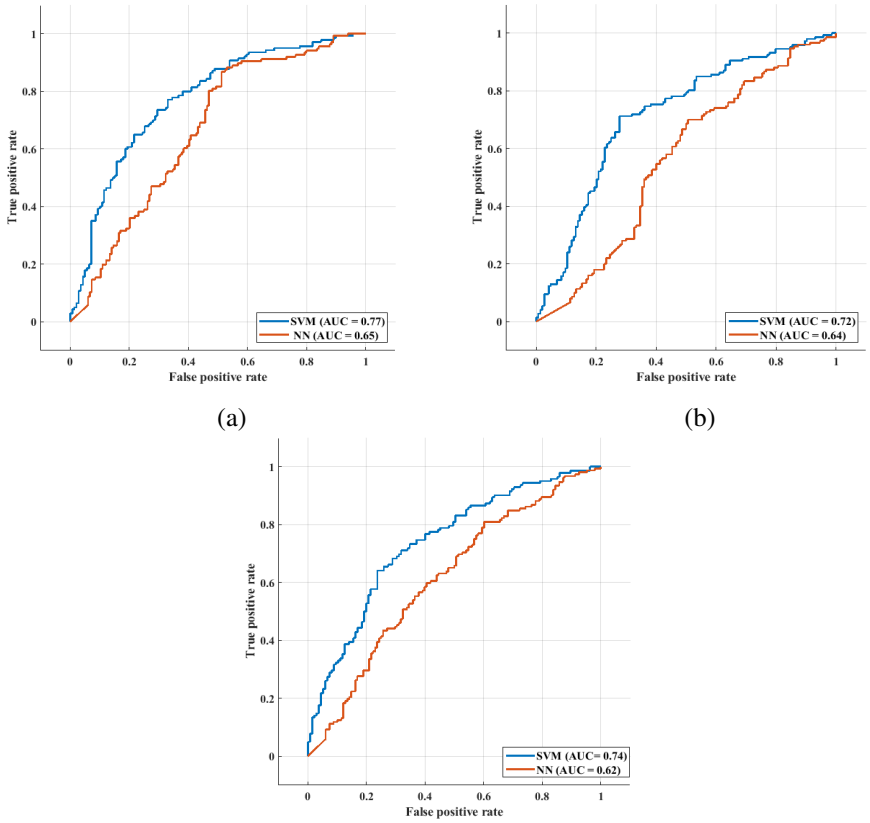


Figure 4.17: ROC curves and AUC values using three features combination models: (a) SI + PPI + TLI, (b) PPI + TLI + PWB, and (c) SI + TLI + PWB , all applied to FSHS dataset

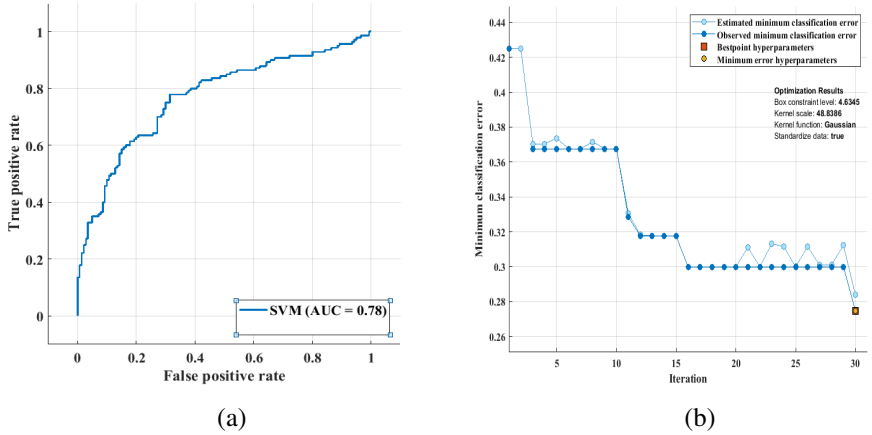
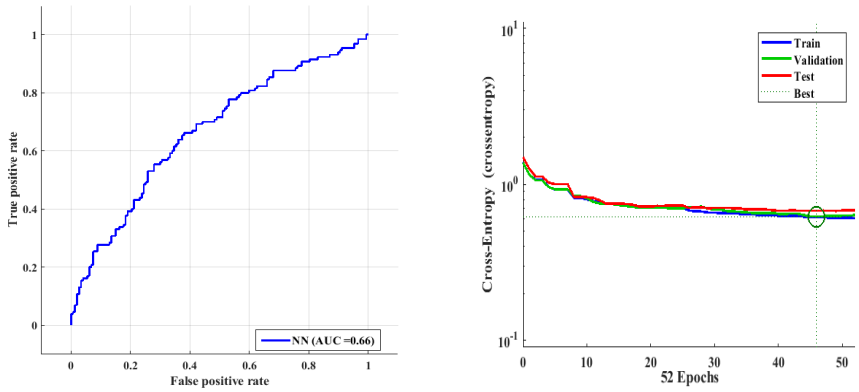


Figure 4.18: Performance of the SVM model using all features combination applied to FSHS dataset in terms of (a) ROC curve, and (b) The minimum classification error



(a) (b)

Figure 4.19: Performance of the NN model using all features combination applied to FSHS dataset in terms of (a) ROC curve, and (b) Loss function

We used the public Khatt dataset to compare the outcomes of our experiments with the state-of-art. The accuracy rates of using the proposed approach on the Khatt dataset are shown in Table 4.11. The results demonstrate that text lines irregularity (TLI) has the highest accuracy among the other individual features, with accuracy rates of 63.8% and 60.9% using SVM and NN, respectively. SI came in second, with accuracy rates of 62.3% and 59.7% using SVM and NN, respectively, while the combination of all features outperformed the state-of-the-art methods with an accuracy rate of 65.2% using SVM method and 67% using NN method.

The AUC-ROC curve of the individual features applied to the Khatt dataset is shown in Figure 4.20. The TLI feature has the highest AUC, with values of 0.67 and 0.62 for SVM and NN, respectively. Using the SVM method, the PPI feature has the lowest AUC value of 0.54.

Figure 4.21 displays the SVM and NN methods' ROC curves and AUC values for the combination of all features. The SVM model's ROC curves with an AUC value of 0.71 are shown in (a). The ROC curve for the NN model with an AUC of 0.63 is shown in (b).

Extracted features				SVM classifier					NN classifier				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (adult)	Mature (Adult)	Youth (adult)	Mature (adult)		Youth (adult)	Mature (adult)	Youth (adult)	Mature (adult)
✓	×	×	×	62.3	60.5	65.4	74.3	50	59.7	51.6	66.7	56.9	61.7
×	✓	×	×	53.6	52.8	54.5	55.9	51.4	50.7	54.1	46.9	54.1	46.9
×	×	✓	×	63.8	61.9	66.7	74.3	52.9	60.9	69.7	52.8	57.5	65.5
×	×	×	✓	59.4	60	58.8	60	58.8	52.2	50.8	60	88.2	20
✓	✓	✓	✓	65.2	62.5	69	73.5	57.1	67	67.8	66.2	65.6	68.5

Table 4.11: Results of the proposed age detection method applied to Khatt dataset

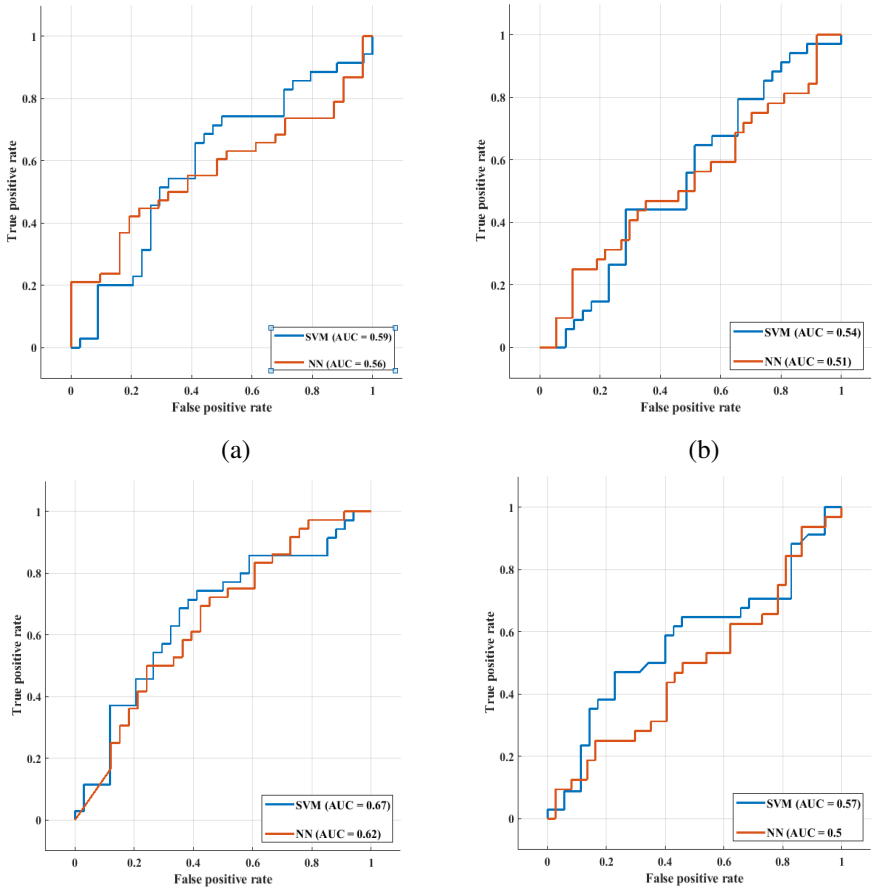


Figure 4.20: ROC curves of SVM and NN results using individual features models: (a) SI, (b) PPI, (c) TLI, and (d) PWB applied on Khatt dataset.

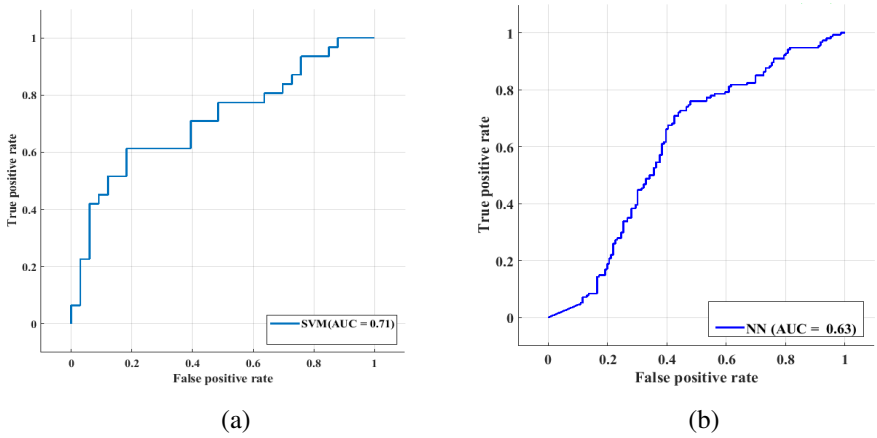


Figure 4.21: ROC curves of SVM and NN results using all features models: (a) SVM and (b) NN, applied to Khatt dataset.

Method	Extracted Features	Classifier	Accuracy (100%)
Bouadjerek et al.	1. Histogram of oriented gradient	SVM	60.3
	2. Gradient local binary Patterns		
Basavaraja et al.	Disconnectedness	K-means	64.44
Proposed Method	**1. Slant irregularity (SI)**	**SVM**	**65.2**
	2. Pen pressure irregularity (PI)		
	3. Text-lines irregularity (ITL)	**NN**	**67**
	4. Percentage of black and white pixels (PWB)		

Table 4.12: Comparison between the proposed age detection method and the available works

Table 4.12 shows a comparison between the proposed method with other available methods. Accuracy, extracted features and the used classifier are presented in the table for better comparison. Similar to the available works on the khatt dataset, we adapted two age ranges, ”16-25” and ”26 and above”. The proposed method outperformed the state-of-the-art with an accuracy rate of 65.2 using a supervised classification method SVM and 67% using NN method similarly to work in [21]. On the other hand, the unsupervised classification method used in [70] has achieved an accuracy rate of 64.44%.

In addition to performing the proposed method on the FSHS dataset and the Khatt dataset, we examined the ability of the proposed features to classify the age groups of handwritings written by only female writers and male writers. So the FSHS dataset was divided into two main subsets: female and male. Then the two age ranges were determined for each sub-set individually. To illustrate, the number of youth adult handwritten samples in the female subgroup were 465, and the number of mature adult handwritten samples were 435. At the same time, there were 435 samples of youth adult male writers and 465 for mature adult male writers.

Table 4.13 shows the results obtained by applying the proposed system on female and male subsets separately. The full feature combinations results are available in B. It can be noticed that the PPI can classify the age of the female writers better than male writers, where it gave an accuracy rate of only 54.5% for male writers compared to 63.6% accuracy rate for female writers. In addition, PPI performs well in classifying samples in the female youth category. In this case, 39.7% of the

test samples were correctly classified as youth female writers, and 23.9% of the test sample were correctly classified as mature writers. This is in contrast to SI, which better classifies mature female writers than youth female writers, as 94% of the correctly classified samples belonged to mature female writers. This also applies to mature male writers, as the SI classified 57.8% of correctly classified samples as mature male writers and 42.2% as youth male writers. It implies that mature writers write more consistently.

In both the subsets of male and female, TLI has the greatest accuracy rate among the other individual feature models. In male handwriting, it produced a classification accuracy of 64.7%, while in female handwriting, it generated an accuracy rate of 66.7%. When the full features were combined, it is clear that it produced the highest accuracy on both female and male subsets, with accuracy rates of 71.9% and 68%, respectively.

The majority of samples that were wrongly classified using individual or combined feature models belonged to the 21-28 age group in both female and male subsets, which is regarded as the border region between the two major classes. Increasing the sample size for each age group to include additional handwriting styles and retraining the models using a new, larger dataset are our recommendations for resolving this issue.

Extracted features				Female Subset					Male Subset				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)		Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)
✓	×	×	×	56.9	50	57.3	7	94.7	62.7	66	60.8	53.1	72.7
×	✓	×	×	63.6	60.7	68.9	78.6	48.2	54.5	54.4	54.5	57.4	51.6
×	×	✓	×	66.7	66.4	66.9	66	67.3	64.7	62.9	66.9	71.3	58
×	×	×	✓	57.3	53.3	53.2	52.4	43.9	53.2	53.3	53.2	52.4	54
✓	✓	✓	✓	71.9	69.4	75.3	79.1	64.7	68	67.7	68.3	68.8	67.2

Table 4.13: Results of the proposed age detection system applied to female and male subsets of FSHS dataset

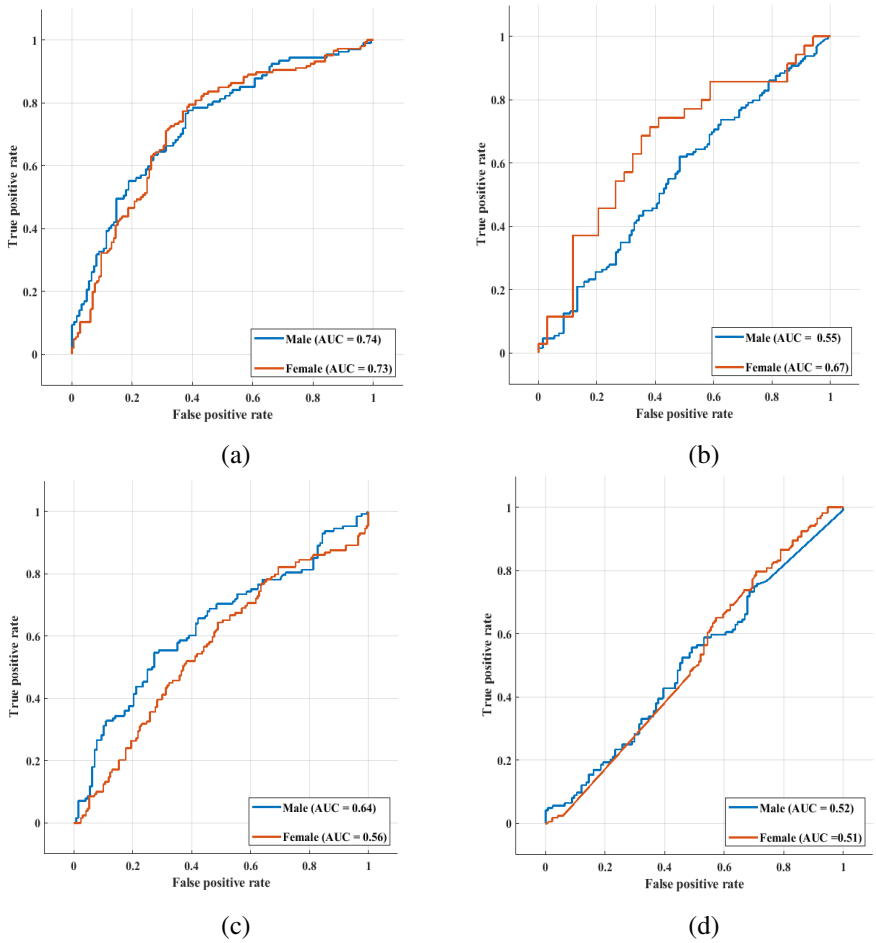
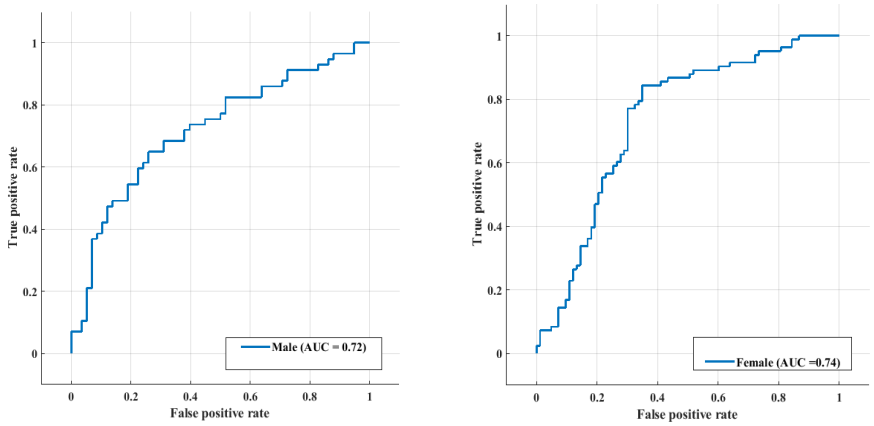


Figure 4.22: ROC curves and AUC results using individual features models: (a) TLI, (b) PPI, (c) SI, and (d) PWB applied on female and male subsets of the FSHS dataset.

The proposed age detection system's performance applied to both female and male subsets of the FSHS dataset was further assessed using the Receiver Operating Characteristic curve (ROC) and the Area Under Curve (AUC).

The ROC curves and AUC results of applying the SVM method to the individual features of female and male subsets of the FSHS dataset are shown in Figure 4.22. For example, the TLI feature is shown in Figure 4.22 (a) for both male and male subsets, with AUC values of 0.74 and 0.73, respectively, the highest among the other features. In contrast, the PWB feature in Figure 4.22 (d) has the lowest AUC values of 0.51 for both male and female subsets. For the individual features, these results are consistent with those provided in Table B.1.



(a) (b)

Figure 4.23: ROC curves and AUC values of the proposed age detection system of ALL features combinations models: (a) Male and (b) Female, applied on female and male subset of the FSHS dataset

The ROC curves and AUC values for the all features combination are shown in Figure 4.23. Figure 4.23 (a) shows the ROC curve for the male subset, which has an AUC of 0.72, while Figure 4.23 (b) shows the ROC curve for the female group, which has an AUC of 0.74.

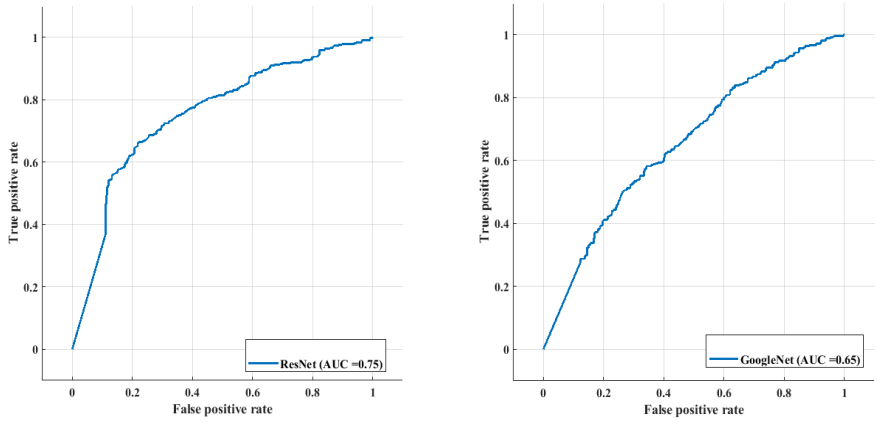
In addition to using traditional image processing techniques to extract the features, we used the transfer learning technique to automatically extract the features from the handwritten documents using two CNN architectures: ResNet and GoogleNet.

Table 4.14 displays the outcomes of applying transfer learning to the FSHS dataset using ResNet and GoogleNet architectures, in terms of accuracy, precision, and recall. ResNet achieved an accuracy of 69.7%, which is higher than GoogleNet's accuracy of 61.1%. Moreover, Figure 4.24 also shows the ROC and AUC values for the two architectures. For example, the ResNet in (a) has a greater AUC value of 0.75 , whereas the GoogleNet in (b) has an AUC value of 0.65.

The transfer learning technique was also applied to two separate subsets of the FSHS dataset: female and male subsets. The results of utilising ResNet and GoogleNet for female and male subsets are shown in Table 4.15. When applied to the male subset, the ResNet had a better accuracy rate

CNN	Accuracy (100%)	Precision (100%)		Recall (100%)	
		Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)
ResNet	69.7	64.8	74.8	72.9	67
GoogleNet	61.1	63.2	58.9	61.6	60.5

Table 4.14: Results of applying the proposed automatic feature extraction method to the FSHS dataset



(a) (b)

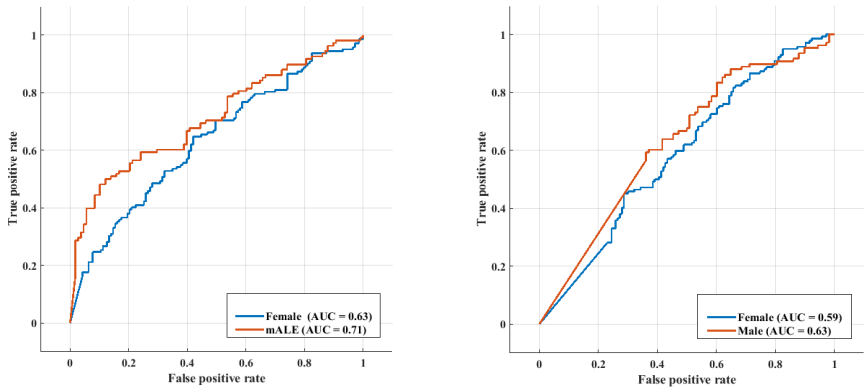
Figure 4.24: ROC curves and AUC values of applying transfer learning to FSHS dataset using (a) ResNet and (b) GoogleNet

CNN	Subsets	Accuracy (100%)	Precision (100%)		Recall (100%)	
			Youth (Adult)	Mature (Adult)	Youth (Adult)	Mature (Adult)
ResNet	Female	60.4	58	62.7	61	59.7
	Male	61.6	55.6	67.6	63.2	60.3
GoogleNet	Female	56.5	51	62	57.5	55.7
	Male	59.7	39.8	79.6	66.2	57

Table 4.15: Results of applying the transfer learning technique to the female and male subsets of FSHS dataset

of 61.6%, compared to 60.4% when applied to the female subset. GoogleNet, on the other hand, appears to be better when applied to the male subset, with an accuracy rate of 59.7%, but only 56.5% when applied to the female subset.

Figure 4.25 shows the ROC curves and AUC values of applying the transfer learning technique to the female and male subsets. The AUC values of 0.63 and 0.71 obtained by applying the ResNet



(a) (b)

Figure 4.25: ROC curves and AUC values of applying Transfer learning using (a) ResNet and (b) GoogleNet to female and male subsets of the FSHS dtataset

to female and male subsets are shown in Figure 4.25 (a). Figure 4.25 (b) depicts the ROC and AUC values obtained from applying GoogleNet on the female subset, with an AUC of 0.59 and an AUC of 0.63 when applied to the male subset.

The experimental findings show that traditional methods of extracting features yield a better classification results than the automatic method based on transfer learning because the second method's extracted features are not directly related to the writer's characteristics.

Chapter 5

Conclusion and Future Works

The goal of this study is to develop and implement automatic gender and age detection systems derived from handwritten documents that can overcome all of all the challenges associated with manual handwritten analysis, which mostly relies on the graphologist's skills.

This research shows that the automatic detection of gender and age from handwritten documents using the computer without human intervention is a viable possibility. Gender detection from handwritten images is a widely discussed topic in a variety of fields, including psychology, document analysis, paleography, graphology, and forensics. At the same time, age detection systems are crucial in the forensic and health care fields.

The success of handwritten classification systems is determined by the discriminality power of the extracted features from handwritten documents, which is a difficult task due to the high degree of similarity between people's handwriting. Extracting the most distinctive features from raw data and representing them in a smaller set of feature vectors is a crucial step. Image processing and transfer learning were utilised in this research to extract features. A set of highly discriminative gender and age-relevant features recommended by a graphologist and a psychologist, such as pen pressure (PP), word spacing (WS), text line irregularity (TLI), and margins (LRM) for gender detection, were extracted using the image processing technique. For age detection, we also took into

account the writing speed and four distinguishing features: slant irregularity (SI), text line irregularity (TLI), pen pressure irregularity (PPI), and proportion of black and white pixels (PWB). We have also developed some effective algorithms for extracting all these features. Chapter 3 contains the details of the extracted features.

We divided our experiments into four categories: individual features models, two features models, three features models and full features models. For both age and gender detection systems, the TLI performed best among the other individual features. In the gender detection system, (WS + TLI) produced the highest accuracy rate of the two feature combinations.

In addition, of the three feature combinations, the combination (LRM + WS + PP) produced the lowest performance. In the age detection system, (SI + PPI + TLI) has the best performance. Finally, among the other feature combinations, the full features combination for both systems has the highest accuracy rate.

We built an Arabic dataset named FSHS to analyse and test the proposed systems (Section 3.1). In addition, we compared our work to other available works using public datasets such as ICDAR2013 and Khatt dataset (Section 4.0.1). When applied to the Arabic subset of the ICDAR2013 dataset, the proposed gender detection system outperforms the state-of-the-art with accuracy rates of 96.2% and 92.5%, using SVM and NN methods, respectively. On the English subset, we attained an accuracy rate of 88.7% using the SVM approach. Furthermore, when applied to the Khatt data set, the proposed age detection system outperforms the state-of-the-art with accuracy of 65.2% and 67%, respectively, using SVM and NN.

Furthermore, female and male subsets of the FSHS dataset were used to test the age detection method. The findings revealed that different features have varying effects on different subsets. To give an example, the female subset responds better to PPI model than the male subset. whereas the male subset has better performance using SI compared to the female subset. In addition, integrating all features gave the best performance for both subsets wit accuracy rates of 71.9% and 68% using

SVM and NN methods, respectively.

Moreover, the transfer learning approach was employed to automatically extract features utilising the knowledge of GoogleNet and ResNet, since the two CNN architectures were used as fixed feature extractors. In the gender detection system, applying the automatic feature extraction method to the FSHS dataset produces accuracy rates of 84.9% and 82.2% using ResNet and GoogleNet, respectively. While the age detection system using the automatic feature extraction method achieved accuracy rates of 69.7% and 61.1% using ResNet and GoogleNet, respectively. We observed that the manual features produced a better classification rate than the automatically extracted features, which is because the manual features are more related to the writer characteristics, whereas the automatic features are more general, such as edges and curves. The experimental results, detailed performance evaluations and analysis are presented in Chapter 4.

Although our framework's findings outperform those of other recently proposed methods in the literature, numerous promising research directions could be explored to improve the proposed work:

- Generalizing the proposed systems so that they could detect age and gender from a variety of languages.
- Exploring further age and gender-related features.
- Investigating more aspects of the writer's characteristics indicated by the handwriting.
- Working on signatures and sketches to reveal the writers' characters.
- Gathering a larger dataset with various labelling that can be used in various research fields, such as stress detection from handwritten documents.
- Expanding the age detection systems to include more age groups.
- Applying the deep learning method to discover the writers' age and gender from the handwritten documents.

5.1 List of Publications

The following are the list of publications that i have worked on through my PhD journey.

- AL-Qawasmeh N., Khayat M. (2020) Automating Stress Detection from Handwritten Documents. In: Lu Y., Vincent N., Yuen P.C., Zheng WS., Cheriet F., Suen C.Y. (eds) Pattern Recognition and Artificial Intelligence. ICPRAI 2020. Lecture Notes in Computer Science,vol12068.Springer,Cham.https://doi.org/10.10 07/978-3-030-59830-313. (published)

- AL-Qawasmeh N., Suen C.Y. (2020) Gender Detection from Handwritten Documents Using Concept of Transfer-Learning. In: Lu Y., VincentN., Yuen P.C., Zheng WS., Cheriet F., Suen C.Y. (eds) Pattern Recognition and Artificial Intelligence. ICPRAI 2020. Lecture Notes in Computer Science, vol 12068. Springer, Cham. https://doi.org/10.1007/978-3-030- 59830-31. (published)

- AL-Qawasmeh N., Khayat M., Suen C.Y. (2022) Novel Feature Extraction methods to Detect Age from Handwriting The 20th Conference of the International Graphon omics Society IGS 2021. (published)

- Najla Alqawasmeh, Ching Y. Suen (2022). Transfer Learning to Detect Age and Gender from Handwritten Documents. Adv. Artif. Intell. Mach. Learn., 2 (2):394-406. (published)

- AL-Qawasmeh Najla, Muna Khayyat, and Ching Y Suen. Novel features to detect gender from handwritten documents. Pattern Recognition Letters, 2022. (published)

- AL-Qawasmeh N., Khayat M., Suen C.Y. (2022) Age Detection from Handwritten Documents. Pattern Recognition letters Journal. (submitted)

Appendix A

Appendix A

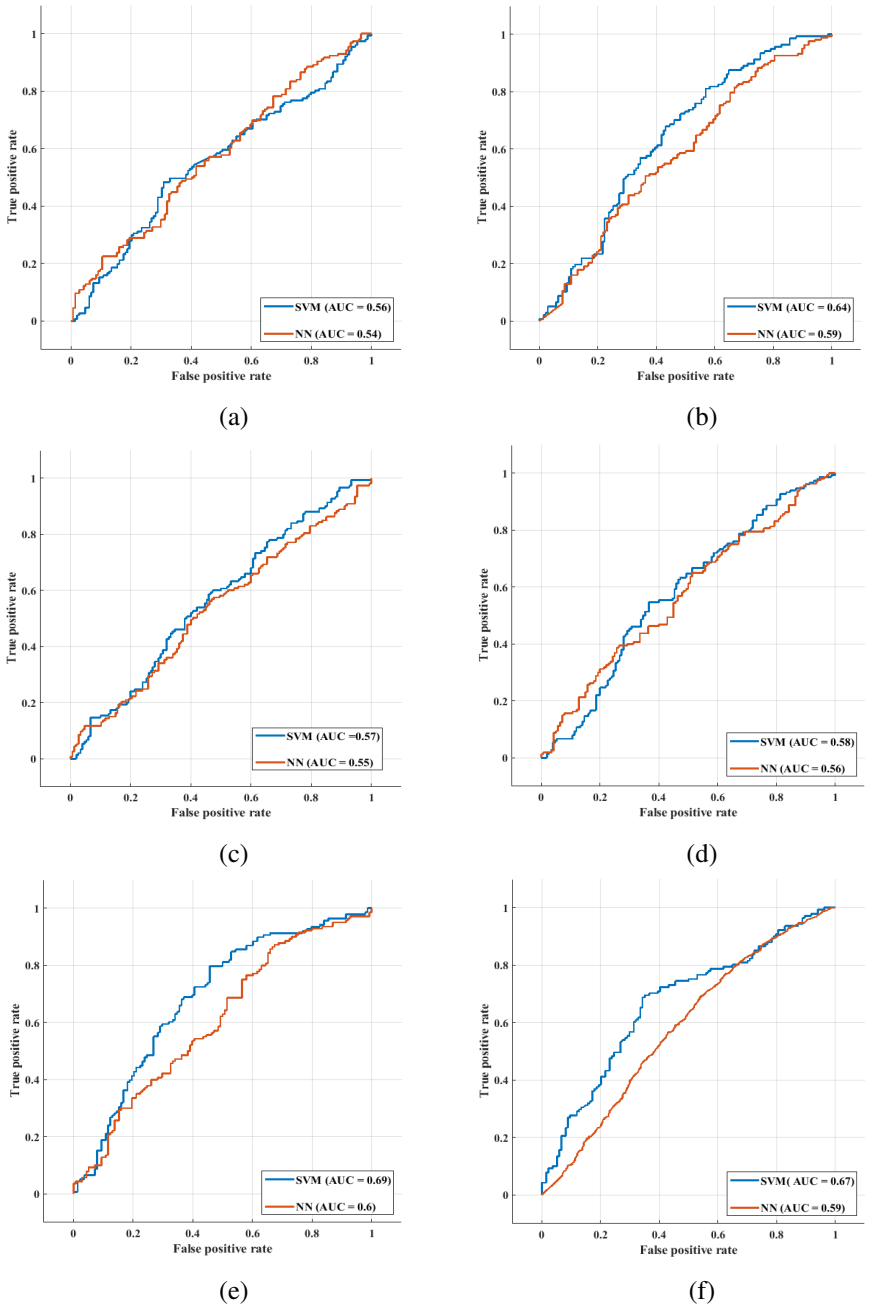


Figure A.1: ROC curves and AUC values using two feature combination models: (a) PPI + PWB, (b) PPI + TLI, (c) SI + PPI, (d) SI + PWB, (e) SI + TLI, and (f) TLI + PWB, all applied to FSHS dataset

Extracted features				SVM classifier					NN classifier				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (adult)	Mature (adult)	Youth (adult)	Mature (adult)		Youth (adult)	Mature (adult)	Youth (adult)	Mature (adult)
✓	×	×	×	56.3	56.1	56,6	58	54.7	56.8	49.2	63.5	54.2	58.8
×	✓	×	×	53.8	54.4	53.4	49.2	58.5	52.7	98.7	6.0	51.6	81.6
×	×	✓	×	66.2	64.7	68.2	71.8	66.7	58.5	44.9	73.1	64.2	55.2
×	×	×	✓	55.7	63.4	53.5	27.5	83.9	55.7	34.6	77.6	61.6	53.3
✓	✓	×	×	55.3	55.3	55.3	55.3	55.3	53.3	46.3	60.1	52.7	53.8
✓	×	✓	×	64.3	62.4	67.2	73.5	55	56.8	60.1	53.6	56.1	57.7
✓	×	×	✓	56.7	56.8	56.5	55.3	58	54.3	63.6	46.3	50.9	59.2
✓	✓	✓	×	70.7	69.1	72.6	75.5	65.8	62.3	62.2	62.5	66.7	57.8
✓	✓	×	✓	60.7	60.8	60.5	61.6	59.7	58	66.5	46.9	62.1	52
✓	×	✓	✓	67.3	66.7	68.1	70.2	64.4	59.2	66.2	52.4	57.5	61.5
×	✓	✓	×	60.3	59.9	60.9	64.2	56.4	56.0	68.1	45.7	51.6	62.7
×	✓	×	✓	54.7	81.3	27.6	50.9	61.4	53.1	54.7	54.7	58.3	51.0
×	×	✓	✓	64.3	62.4	66.6	72	56.7	57.0	55	59.2	60.7	53.5
×	✓	✓	✓	69.7	68.8	70.6	72	67.3	60	64.0	56.0	59.3	61
✓	✓	✓	✓	71	68.9	73.7	76.7	65.3	63.5	64.6	62.3	66	60.9

Table A.1: Results of applying the proposed age detection system to FSHS dataset

Appendix B

My Appendix

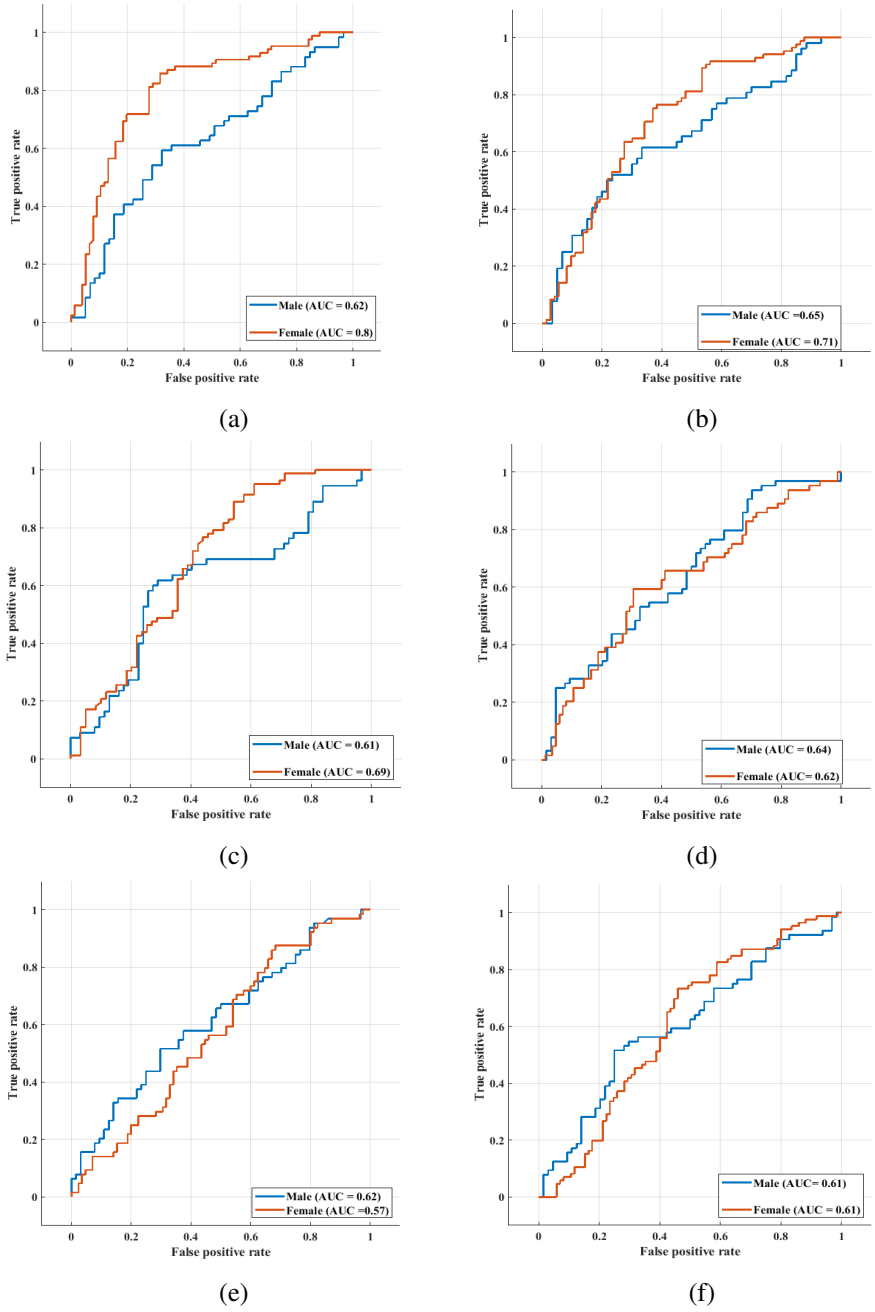


Figure B.1: ROC curves and AUC values of the proposed age detection system of two feature combination models: (a) TLI + PWB, (b) PPI + TLI, (c) SI + TLI (d) SI + PPI, (e) SI + PWB, and (f) PPI + PWB, applied on female and male subset of the FSHS dataset

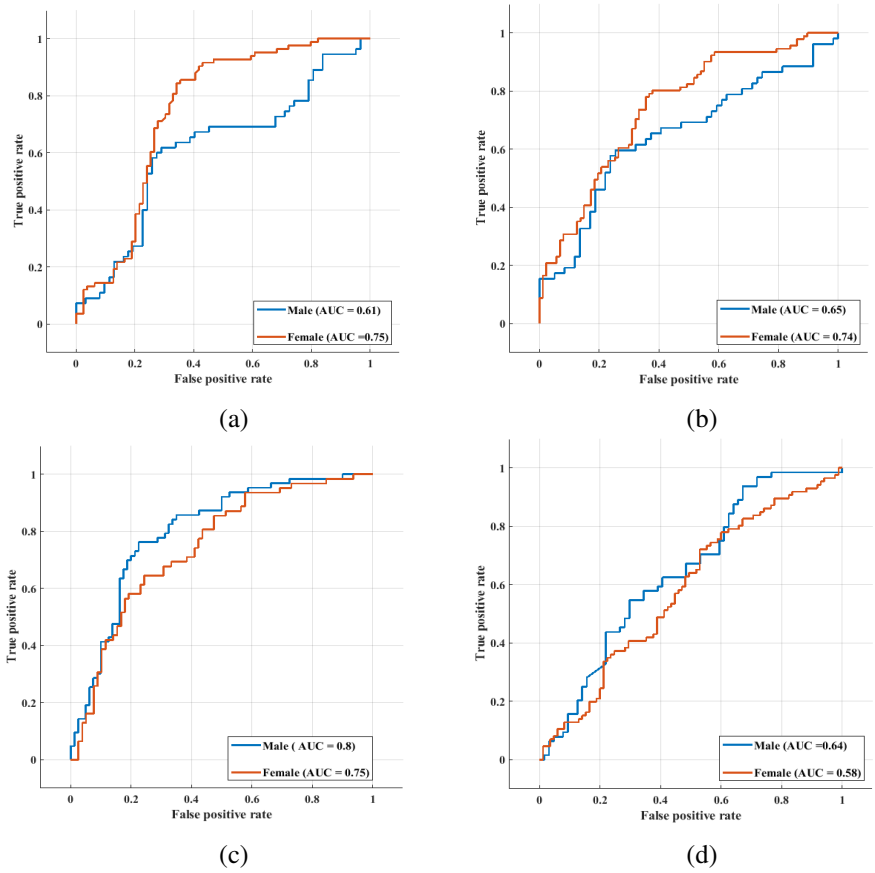


Figure B.2: ROC curves and AUC values of the proposed age detection system of three feature combination models: (a) SI + TLI + PWB, (b) PPI + TLI + PWB, (c) SI + PPI + TLI, and (d) SI + PPI + PWB, applied on female and male subset of the FSHS dataset

Extracted features				Female subset of FSHS dataset					Male subset of FSHS dataset				
SI	PPI	TLI	PWB	Accuracy (100%)	Precision (100%)		Recall (100%)		Accuracy (100%)	Precision (100%)		Recall (100%)	
					Youth (adult)	Mature (adult)	Youth (adult)	Mature (adult)		Youth (adult)	Mature (adult)	Youth (adult)	Mature (adult)
✓	×	×	×	56.9	50	57.3	7	94.7	62.9	66	60.8	53.1	72.7
×	✓	×	×	63.6	60.7	68.9	78.6	48.2	54.5	54.4	54.5	57.4	51.6
×	×	✓	×	66.7	66.4	66.9	66	67.3	64.7	62.9	66.9	71.3	58
×	×	×	✓	53.6	53.2	54.3	63.4	43.9	53.2	53.3	53.2	52.4	54
✓	✓	×	×	59.7	54.3	62.1	39.1	75.3	59.4	60.3	58.6	54.7	64.1
✓	×	✓	×	62.4	60.6	100	100	11.1	67.2	62.2	67.2	67.2	67.2
✓	×	×	✓	58.4	60	58.3	9.4	95.3	59.4	60	58.8	56.3	62.5
✓	✓	✓	×	69.8	67.3	71.3	57.8	78.8	73.8	74.5	73.5	59.4	84.7
✓	✓	×	✓	57.3	56.8	57.9	62.8	51.8	60.2	55.8	93.3	98.4	21.9
✓	×	✓	✓	73.1	68.5	81	86	60	67.2	67.2	67.2	67.2	67.2
×	✓	✓	×	62.6	60	67.2	76.7	48.2	64.1	64.1	64.1	64.1	64.1
×	✓	×	✓	60.9	62.1	60	56.3	65.6	61.7	64.2	60	53.1	70.3
×	×	✓	✓	72.5	69.7	76.4	80.2	64.7	60.9	62.1	60	56.3	65.6
×	✓	✓	✓	69.4	66.7	73.3	78.3	60.4	66.4	66.2	66.7	67.2	65.6
✓	✓	✓	✓	71.9	69.4	75.3	79.1	64.7	68	67.7	68.3	68.8	67.2

Table B.1: Results of the proposed age detection system applied to female and male subsets of FSHS dataset

Bibliography

[1] Seema Kedar, Vaishnavi Nair, and Shweta Kulkarni. Personality identification through handwriting analysis: a review. *Int. J. Adv. Res. Comput. Sci. Softw. Eng*, 5(1), 2015.

[2] Baruch Nevo. Validation of graphology through use of a matching method based on ranking. *Perceptual and Motor Skills*, 69(3_suppl):1331–1336, 1989.

[3] Mohd Razif Shamsuddin, Ku Shairah, Zaidah Ibrahim, Rahmattullah Khan, and Azlinah Mohamed. Graphology and cattell's 16pf traits matrix (holocatt matrix). *Convergence Information Technology, International Conference on*, 1:220–225, 11 2008.

[4] Oscar Miguel-Hurtado, Richard Guest, Sarah Stevenage, and Greg Neil. The relationship between handwritten signature production and personality traits. *IJCB 2014- IEEE/IAPR International Joint Conference on Biometrics*, pages 1–8, 09 2014.

[5] H. N. Champa and K. Anandakumar. Automated human behavior prediction through handwriting analysis. *2010 First International Conference on Integrated Intelligent Computing*, pages 160–165, 2010.

[6] A. Priya, S. Mishra, S. Raj, S. Mandal, and S. Datta. Online and offline character recognition: A survey. In *2016 International Conference on Communication and Signal Processing (ICCSP)*, pages 0967–0970, 2016.

[7] Yan Solihin and C.G. Leedham. Noise and background removal from handwriting images. In *Proceedings Intelligent Information Systems. IIS'97*, pages 366–370, 1997.

[8] Atena Farahmand, Abdolhossein Sarrafzadeh, and Jamshid Shanbezadeh. Document image noises and removal methods. *Lecture Notes in Engineering and Computer Science*, 2202:436–440, 2013.

[9] Atallah M.AL-Shatnawi and Khairuddin Omar. The thinning problem in arabic text recognition a comprehensive review. *International Journal of Computer Applications*, 103:35–42, 2014.

[10] Hanan Aljuaid, Zulkifli Muhammad, and Muhammad Sarfraz. A tool to develop arabic handwriting recognition system using genetic approach. *Journal of Computer Science*, 6(6):619–624, 2010.

[11] Muhannad Aulama, Asem Natsheh, Gheith Abandah, and Mohammed Olama. Optical character recognition of handwritten arabic using hidden markov models. *Proceedings of SPIE - The International Society for Optical Engineering*, 8055, 2011.

[12] Atallah Al-Shatnawi, Safwan Al-Salaimeh, and Farah Hanna Al-Zawaideh. Offline arabic text recognition – an overview. 2011.

[13] Enric Sesa-Nogueras, Marcos Faundez-Zanuy, and Josep Roure-Alcobé. Gender classification by means of online uppercase handwriting: A text-dependent allographic approach. *Cognitive Computation*, 8(1):15–29, 2016.

[14] Ali Mirza, Momina Moetesum, Imran Siddiqi, and Chawki Djeddi. Gender classification from offline handwriting images using textural features. In *2016 15th International Conference on Frontiers in Handwriting Recognition*, pages 395–398, 2016.

[15] Chawki Djeddi, Abdeljalil Gattal, Labiba Souici-Meslati, and Imran Siddiqi. Lamis-mshd: A multi-script offline handwriting database. In *2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 93–97, 2014.

[16] Sarah Hamid and Kate Miriam Loewenthal. Inferring gender from handwriting in urdu and english. *The Journal of Social Psychology*, 136(6):778–782, 1996.

[17] James Hartley. Sex differences in handwriting: a comment on spear. *British Educational Research Journal*, 17(2):141–145, 1991.

[18] Prasenjit Maji, Souvik Chatterjee, Sayan Chakraborty, Noreen Kausar, Nilanjan Dey, and Sourav Samanta. Effect of euler number as a feature in gender recognition system from offline handwritten signature using neural networks. In *2nd International Conference on Computing for Sustainable Global Development. INDIACom-2015, At Bharati Vidyapeeth, New Delhi (INDIA)*, pages 1869–1873, 2015.

[19] Payal Maken and Abhishek Gupta. A study on various techniques involved in gender prediction system: A comprehensive review. *Cybernetics and Information Technologies*, 19:51–73, 2019.

[20] A. H. Garoot, M. Safar, and C. Y. Suen. A comprehensive survey on handwriting and computerized graphology. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pages 621–626, Los Alamitos, CA, USA, 2017.

[21] Nesrine Bouadjenek, Hassiba Nemmour, and Youcef Chibani. Age, gender and handedness prediction from handwriting using gradient features. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1116–1120. IEEE, 2015.

[22] Liana Lorigo and Venu Govindaraju. Offline arabic handwriting recognition: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28:712 – 724, 2006.

[23] G. Kumar and P. K. Bhatia. A detailed review of feature extraction in image processing systems. In *2014 Fourth International Conference on Advanced Computing Communication Technologies*, pages 5–12, 2014.

[24] Pierre Michel Lallican, Christian Viard-Gaudin, and Stefan Knerr. From off-line to on-line handwriting recognition. In *Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition*, pages 303–312, 2000.

[25] Marcus Liwicki and Horst Bunke. Iam-ondb - an on-line english sentence database acquired from handwritten text on a whiteboard. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pages 956–961 Vol. 2, 2005.

[26] Cheng-Lin Liu, Fei Yin, Da-Han Wang, and Qiu-Feng Wang. Casia online and offline chinese handwriting databases. pages 37 – 41, 2011.

[27] Fouad Slimane, Slim Kanoun, Haikal El Abed, Adel M. Alimi, Rolf Ingold, and Jean Hennebert. Icdar2013 competition on multi-font and multi-size digitally represented arabic text. *2013 12th International Conference on Document Analysis and Recognition*, pages 1433–1437, 2013.

[28] Shikha Magotra, Baijnath Kaushik, and Ajay Kaul. A database for printed takri class of northwest indian regional scripts. In Pradeep Kumar Singh, Sanjay Sood, Yugal Kumar, Marcin Paprzycki, Anton Pljonkin, and Wei-Chiang Hong, editors, *Futuristic Trends in Networks and Computing Technologies*, pages 508–520, Singapore, 2020.

[29] Hamzah Luqman and Mohammed Assayony. Using bag-of-features for arabic font recognition. pages 628–632, 2021.

[30] Jaroslaw Moszczynski. The multi-individuality of handwriting. *Forensic Science International*, 294:e4–e10, 2019.

[31] Mario Pechwitz, Samia Snoussi Maddouri, Volker Märgner, Noureddine Ellouze, and Hamid Amiri. Ifn/enit - database of handwritten arabic words. In *In Proc. of CIFED 2002*, pages 129–136, 2002.

[32] U.-V. Marti and H. Bunke. The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[33] Jabril Ramdan, Khairuddin Omar, Mohammad Faidzul, and Ali Mady. Arabic handwriting data base for text recognition. *Procedia Technology*, 11:580 – 584, 2013. 4th International Conference on Electrical Engineering and Informatics, ICEEI 2013.

[34] Emmanuèle Grosicki, Matthieu Carré, J.-M. Brodin, and Edouard Geoffrois. Results of the rimes evaluation campaign for handwritten mail processing. *2009 10th International Conference on Document Analysis and Recognition*, pages 941–945, 2009.

[35] Stephanie Strassel. Linguistic resources for arabic handwriting recognition. In Khalid Choukri and Bente Maegaard, editors, *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, Cairo, Egypt, April 2009. The MEDAR Consortium.

[36] Somaya Al-ma'adeed, Wael Ayouby, Abdelâali Hassaïne, and Mahallal Jihad. Quwi: An arabic and english handwriting dataset for offline writer identification. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, pages 746–751, 2012.

[37] Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In *2013 12th International Conference on Document Analysis and Recognition*, pages 560–564, 2013.

[38] Abdelâali Hassaïne, Somaya Al Maadeed, Jihad Aljaam, and Ali Jaoua. Icdar 2013 competition on gender prediction from handwriting. In *12th International Conference on Document Analysis and Recognition*, pages 1417–1421, 2013.

[39] Chawki Djeddi, Abdeljalil Gattal, Labiba Souici-Meslati, Imran Siddiqi, Youcef Chibani, and Haikal El Abed. Lamis-mshd: A multi-script offline handwriting database. In *In The 14th International Conference on Frontiers in Handwriting Recognition, ICFHR, Crete, Greece, September 1-4*, pages 93–97. IEEE, 2014.

[40] Sabri A. Mahmoud, Irfan Ahmad, Wasfi G. Al-Khatib, Mohammad Alshayeb, Mohammad Tanvir Parvez, Volker Märgner, and Gernot A. Fink". Khatt: An open arabic offline handwritten text database. *Pattern Recognition*, 47(3):1096 – 1112, 2014.

[41] Imen Chtourou, Ahmed Cheikh Rouhou, Faten Kallel Jaiem, and Slim Kanoun. Altid : Arabic/latin text images database for recognition research. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 836–840, 2015.

[42] Abdelaali Hassaïne, Somaya Al-Maadeed, Jihad Mohamad Alja'am, Ali Jaoua, and Ahmed Bouridane. The icdar2011 arabic writer identification contest. In *2011 International Conference on Document Analysis and Recognition*, pages 1470–1474, 2011.

[43] Alicia Fornés, Anjan Dutta, Albert Gordo, and Josep Lladós. Cvc-muscima: a ground truth of handwritten music score images for writer identification and staff removal. *International Journal on Document Analysis and Recognition (IJDAR)*, 15:243–251, 2011.

[44] Huaigu Cao, Rohit Prasad, and Prem Natarajan. Improvements in hmm adaptation for handwriting recognition using writer identification and duration adaptation. In *2010 12th International Conference on Frontiers in Handwriting Recognition*, pages 154–159, 2010.

[45] Nayef N. Kharma, Mahmoud Ali Ahmed, and Rod. Ward. A new comprehensive database of handwritten arabic words, numbers, and signatures used for ocr testing. *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No.99TH8411)*, 2:766–768 vol.2, 1999.

[46] Charilaos Akasiadis and Ioannis A. Vetsikas. A dataset generator for evaluating resource allocation algorithms. 2015.

[47] Najla AL-Qawasmeh and Ching Y. Suen. Gender detection from handwritten documents using concept of transfer-learning. In Yue Lu, Nicole Vincent, Pong Chi Yuen, Wei-Shi Zheng, Farida Cheriet, and Ching Y. Suen, editors, *Pattern Recognition and Artificial Intelligence*, pages 3–13, Springer, 2020.

[48] R.A. Huber and A.M. Headrick. *Handwriting Identification: Facts and Fundamentals*. CRC Press, 1999.

[49] Abdeljalil Gattal, Chawki Djeddi, Imran Siddiqi, and Youcef Chibani. Gender classification from offline multi-script handwriting images using oriented basic image features (obifs). *Expert Systems with Applications*, 99:155 – 167, 2018.

[50] Somaya Al Maadeed and Abdelaali Hassaine. Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing*, 2014(1):10, 2014.

[51] Ning Bi, Ching Y Suen, Nicola Nobile, and Jun Tan. A multi-feature selection approach for gender identification of handwriting based on kernel mutual information. *Pattern Recognition Letters*, 121:123 – 132, 2019.

[52] Imran Siddiqi, Chawki Djeddi, Ahsen Raza, and Labiba Souici-Meslati. Automatic analysis of handwriting for gender classification. *Pattern Analysis and Applications*, 18:887–899, 2014.

[53] Payal Maken and Abhishek Gupta. A method for automatic classification of gender based on text-independent handwriting. *Multimedia Tools and Applications*, pages 1–30, 2021.

[54] Shaveta Dargan and Munish Kumar. Gender classification and writer identification system based on handwriting in gurumukhi script. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 388–393, 2021.

[55] Y. Akbari, Kazem Nouri, Javad Sadri, Chawki Djeddi, and Imran Siddiqi. Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. *Image and Vision Computing*, 59:17–30, 2016.

[56] Shivanand S. Gornale, Sathish Kumar, Abhijit Patil, and Prakash S. Hiremath. Behavioral biometric data analysis for gender classification using feature fusion and machine learning. *Frontiers in Robotics and AI*, 8, 2021.

[57] Fahimeh Alaei and Alireza Alaei. Gender detection based on spatial pyramid matching. In Josep Lladós, Daniel Lopresti, and Seiichi Uchida, editors, *Document Analysis and Recognition – ICDAR 2021*, pages 305–317, 2021.

[58] Faycel Abbas, Abdeljalil Gattal, Mohamed Ridda Laouar, Kamel Saoudi, and Ismail Hadjadj. Local binary patterns for gender classification. *Proceedings of the 7th International Conference on Software Engineering and New Technologies*, pages 1–5, 2018.

[59] Evyatar Illouz, Eli (Omid) David, and Nathan S. Netanyahu. Handwriting-based gender classification using end-to-end deep neural networks. In Věra Kůrková, Yannis Manolopoulos, Barbara Hammer, Lazaros Iliadis, and Ilias Maglogiannis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2018*, pages "613–621", 2018.

[60] Angel Morera, Ángel Sánchez, Jose Vélez, and Ana Moreno. Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, 2018:1–14, 2018.

[61] Gang Xue, Shifeng Liu, Daqing Gong, and Yicao Ma. Atp-densenet: a hybrid deep learning-based gender identification of handwriting. *Neural Computing and Applications*, 33(10):4611–4622, 2021.

[62] Irina Rabaev, Marina Litvak, Sean Asulin, and Or Haim Tabibi. Automatic gender classification from handwritten images: A case study. In Nicolas Tsapatsoulis, Andreas Panayides, Theo Theocharides, Andreas Lanitis, Constantinos Pattichis, and Mario Vento, editors, *Computer Analysis of Images and Patterns*, pages 329–339, 2021.

[63] Amira E. Youssef, Ahmed S. Ibrahim, and A. Lynn Abbott. Automated gender identification for arabic and english handwriting. In *5th International Conference on Imaging for Crime Detection and Prevention, ICDP, London, UK.*, pages 1–6. IET / IEEE, 2013.

[64] Chawki Djeddi, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Labiba Souici-Meslati, and Haikal El Abed. Icdar competition on multi-script writer identification and gender classification using 'quwi' database. In *13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1191–1195, 2015.

[65] Himadri Mukherjee, Chandrima Majumder, Suparna Saha Biswas, and Ankita Dhar. A cnn-based approach towards gender identification from online handwritten bangla vowel modifiers. In Jyotsna Kumar Mandal, Rajkumar Buyya, and Debashis De, editors, *Proceedings of International Conference on Advanced Computing Applications*, pages 483–491, Singapore, 2022.

[66] B. Navya, Palaiahnakote Shivakumara, G.C Shwetha, Sangheeta Roy, Devanur Guru, Umapada Pal, and Tong Lu. Adaptive multi-gradient kernels for handwritting based gender identification. In *The 16th International Conference on Frontiers in Handwriting Recognition*, pages 392–397, 2018.

[67] Abdeljalil Gattal, Chawki Djeddi, Ameur Bensefia, and Abdellatif Ennaji. Handwriting based gender classification using cold and hinge features. In Abderrahim El Moataz, Driss Mammass, Alamin Mansouri, and Fathallah Nouboud, editors, *Image and Signal Processing*, pages 233–242, 2020.

[68] G. Louloudis, Basilis Gatos, and N. Stamatopoulos. Icfhr competition on writer identification challenge 1: Latin/greek documents. In *2012 International Conference on Frontiers in Handwriting Recognition*, pages 829–834, 2012.

[69] Chawki Djeddi, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Abdellatif Ennaji, and Haikal El Abed. Icfhr2016 competition on multi-script writer demographics classification using "quwi" database. In *15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 602–606, 2016.

[70] V. Basavaraja, Palaiahnakote Shivakumara, D. S. Guru, Umapada Pal, Tong Lu, and Michael Blumenstein. Age estimation using disconnectedness features in handwriting. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1131–1136, 2019.

[71] Anith Hasseim, Rubita Sudirman, and Puspa Inayat Khalid. Handwriting classification based on support vector machine with cross validation. *Engineering*, 05:84–87, 2013.

[72] Abhishek Bal and Rajib Saha. An improved method for handwritten document analysis using segmentation, baseline recognition and writing pressure detection. *Procedia Computer Science*, 93:403–415, 09 2016.

[73] D. Liu and J. Yu. Otsu method and k-means. In *2009 Ninth International Conference on Hybrid Intelligent Systems*, volume 1, pages 344–349, 2009.

[74] Muna Khayyat, Louisa Lam, Ching Y. Suen, Fei Yin, and Cheng-Lin Liu. Arabic handwritten text line extraction by applying an adaptive mask to morphological dilation. In *10th IAPR International Workshop on Document Analysis Systems, DAS 2012, Gold Coast, Queenslands, Australia, March 27-29, 2012*, pages 100–104, 2012.

[75] Lifeng He, Yuyan Chao, Kenji Suzuki, and Kesheng Wu. Fast connected-component labeling. *Pattern Recognition*, 42(9):1977 – 1987, 2009.

[76] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., USA, 2nd edition, pages 334-347, 2001.

[77] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56(1):3–69, 2014.

[78] Wesley Chin, Alam Harvey, and Andrew Jennings. Skew detection in handwritten scripts. In *TENCON'97 Brisbane-Australia. Proceedings of IEEE TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications (Cat. No. 97CH36162)*, volume 1, pages 319–322. IEEE, 1997.

[79] Syed Moshfeq Salaken, Abbas Khosravi, Thanh Nguyen, and Saeid Nahavandi. Extreme learning machine based transfer learning algorithms: A survey. *Neurocomputing*, 267:516 – 524, 2017.

[80] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In Věra Kůrková, Yannis Manolopoulos, Barbara Hammer, Lazaros Iliadis, and Ilias Maglogiannis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2018*, pages 270–279, 2018.

[81] Harsh Jain. What is transfer learning and why is it needed? https://www.educative.io/edpresso/what-is-transfer-learning-and-why-is-it-needed, 2022. Accessed: 2022-03-08.

[82] Lorien Y. Pratt, Jack Mostow, and Candace A. Kamm. Direct transfer of learned information among neural networks. In *Proceedings of the Ninth National Conference on Artificial Intelligence - Volume 2*, AAAI'91, page 584–589, Anaheim, California, 1991.

[83] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, Los Alamitos, CA, USA, jun 2015.

[84] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[85] Kuldeep Shiruru. An introduction to artificial neural network. *International Journal of Advance Research and Innovative Ideas in Education*, 1:27–30, 09 2016.

[86] M Ikonomakis, Sotiris Kotsiantis, and V Tampakas. Text classification using machine learning techniques. *WSEAS transactions on computers*, 4(8):966–974, 2005.

[87] Xinyang Deng, Qi Liu, Yong Deng, and Sankaran Mahadevan. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340-341:250–261, 2016.

[88] David J Hand. Measuring classifier performance: a coherent alternative to the area under the roc curve. *Machine learning*, 77:103–123, 2009.