AGE AND GENDER PREDICTION FROM 3D-BODY AND FACE IMAGES

A DOCTOR OF PHILOSOPHY THESIS

in

Software Engineering

Atılım University

by

SEDA ÇAMALAN

JUNE 2018

AGE AND GENDER PREDICTION FROM 3D-BODY AND FACE IMAGES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

SEDA ÇAMALAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

JUNE 2018

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

_____

Prof. Dr. Ali Kara

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

_____

Prof. Dr. Ali Yazıcı

Head of Department

This is to certify that we have read the thesis "Age and Gender Prediction from 3D-Body and Face Images" submitted by "Seda Çamalan" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

_____

Assoc. Prof. Dr. Gökhan ŞENGÜL

Supervisor

Examining Committee Members

Assoc. Prof. Dr. Gökhan ŞENGÜL _____

Asst. Prof. Dr. Atila BOSTAN _____

Asst. Prof Dr. Kasım ÖZTOPRAK _____

Asst. Prof Dr. Çiğdem TURHAN _____

Asst. Prof Dr. Erol ÖZÇELİK _____

Date: June 22, 2018

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Seda Çamalan

Signature:

ABSTRACT

AGE AND GENDER PREDICTION FROM 3D-BODY AND FACE IMAGES

ÇAMALAN, Seda

Ph.D., Software Engineering Department

Supervisor: Assoc. Prof. Dr. Gökhan ŞENGÜL

June 2018, 119 pages

The biometric data collected from individuals provide an array of information about any population and their environment which can be used in several areas, including transportation (busses, ferries, railways, etc), shopping malls, public areas, sports centers, museums, supermarkets, libraries, etc., not to mention security applications. In detail, this biometric data is related with identity, gender, race, height, weight, and eye and hair color of the person. In this thesis, an image processing-based system to predict the two major aspects, age range and genders of people is developed and integrated as a software tool. A standard RGB camera is used to acquire face images, while a 3D camera is used for body information. To predict the gender and age of each individual, statistical pattern recognition algorithms, deep learning and neural network-based approaches are utilized. For statistical methods, LBP and HOG methods are applied on face images to extract features, then KNN and SVM classification methods are applied as classifiers. Convolutional neural network is used to predict age range of people and the comparison between statistical methods and convolutional neural networks are presented. For age prediction, from face images, statistical methods results yielding a top accuracy of 40.1%; whereas, the best accuracy obtained from CNN deep learning is 59.1%. In addition, 3D body information is used for gender and age prediction by applying statistical and neural network methods. These methods show to improve the gender prediction rate by up to 99.26% and age

prediction by 99.41% for the whole-body information. The upper-body and lower-body parts are also examined separately to predict the age and gender of the each individual.

Key Words: Age prediction, gender prediction, pattern recognition, neural network, deep learning

ÖZ

3 BOYUTLU VÜCUT VE YÜZ GÖRÜNTÜLERİNDEN YAŞ VE CİNSİYET TAHMİNİ

ÇAMALAN, Seda

Doktora, Yazılım Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Gökhan ŞENGÜL

Haziran 2018, 119 sayfa

İnsanlardan elde edilen biyometrik veriler, insanlar ve çevre hakkında birçok bilgi sağlar. Bu bilgi ulaşım alanları (otobüs, vapur, demiryolu, vb), alışveriş merkezleri, kamu alanları, spor merkezleri, müzeler, süpermarketler, kütüphaneler, vb. gibi birçok alanda kullanılabilir. Birçok alanda dikkate alınan biyometrik veriler cinsiyet, ırk, boy, kilo, göz ve saç rengidir. Bu tez çalışmasında, insanların biyometrik verilerinden yaş aralığını ve cinsiyetlerini tahmin eden bir görüntü işleme tabanlı kombine sistem geliştirilmiş ve bir yazılım aracı haline getirilmiştir. Yüz görüntülerini elde etmek için standart RGB kamera kullanılırken vücut bilgilerini elde etmek için 3D kamera kullanılmaktadır. İnsanların cinsiyet ve yaşını tahmin etmek için istatistiksel örüntü tanıma algoritmaları, derin öğrenme ve yapay sinir ağı tabanlı yaklaşımlar kullanılmıştır. İstatistiki metotlar olarak, LBP ve HOG metotları, özniteliklerin elde edilmesi için yüz görüntülerine uygulanmakta, daha sonra KNN ve SVM sınıflandırıcılar, cinsiyet ve yaş tahmini için kullanılmaktadır. İnsanların yaşını tahmin etmek için yapay sinir ağı da kullanılmıştır ve istatistiksel yöntemler ile yapay sinir ağları arasındaki karşılaştırmalar yapılmıştır. Yaş aralığı tahmini için yüz görüntülerinden istatistiksel yöntemler ile en iyi doğruluk %40,1 olarak elde edilmiştir. CNN derin öğrenmelerinden elde edilen en iyi doğruluk oranı ise %59.1'dir. Yaş ve cinsiyet tahmini için 3D vücut bilgisi de kullanılmıştır. Yapay sinir ağları ile

3D vücut bilgilerinin sınıflandırılması sonucu cinsiyet tahmini başarımı oranını %99,26'ya ve yaş tahmini % 99.41'e yükseltilmiştir. Üst vücut ve alt vücut kısımlarının da insanların yaşının ve cinsiyetininin tahmini için kullanılabileceği değerlendirilmiş ve deneysel çalışmalar yapılmıştır.

Anahtar Kelimeler: Yaş tahmini, cinsiyet tahmini, örüntü tanıma, yapay sinir ağı, derin öğrenme

To My Family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere appreciation to my supervisor Assoc. Prof. Dr. Gökhan ŞENGÜL, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. I learned a lot from him. I offer sincere thanks for his extensive knowledge, enthusiasm and patience during this period.

I would also like to thank Asst. Prof. Dr. Atila BOSTAN, Asst. Prof. Dr. Kasım ÖZTOPRAK, Asst. Prof. Dr. Çiğdem TURHAN and Asst. Prof. Dr. Erol ÖZÇELİK for their valuable time, comments and suggestions, all helping to improve my thesis workcomments on my study.

I would also like to thank RA. Buğra Yener ŞAHİNOĞLU, for his technical advice, supports and valuable time on implementation of codes in C#. Not only the technical support, but also the friendly support and motivation.Additionally, I would like to thank my friends; Damla TOPALLI, for her valuable support, motivation, valuable time and contributions.

I would also like to extend my thanks to the equipment in Atilim University Simulation Laboratory, which is constituted for Endo-neurosurgery Education (ECE: Tubitak 1001, Project No: 112K287) project supported by TÜBİTAK 1001, established by Assoc. Prof. Dr. Nergiz ERCİL ÇAĞILTAY.

Last but not least, I wish to express my appreciation to my dearest parents Senem ÇAMALAN, Ayhan ÇAMALAN and my grandmother Hatun ÇAMALAN for their support, encouragement and loved me for a lifetime.

TABLE OF CONTENTS

ABSTRACT iii
ÖZ v
ACKNOWLEDGMENTS viii
TABLE OF CONTENTS ix
LIST OF TABLES xii
LIST OF FIGURES xiv
LIST OF ABBREVIATIONS xvii
CHAPTER 1 1
INTRODUCTION 1
1.1 Motivation 2
1.2 Research Problem 3
1.3 Outline 4
CHAPTER 2 6
LITERATURE REVIEW 6
2.1 Gender Prediction 6
2.1.1 Gender Prediction from Face Images 7
2.1.2 Gender Prediction from Body Information 9
2.2 Age Prediction 11
CHAPTER 3 16
MATERIALS AND METHODOS 16
3.1 Proposed Gender and Age Prediction System 17
3.1.1 Data Collection 17
3.1.2 Gender Prediction 19
3.1.3 Age Prediction 20
3.1.4 Graphical User Interface 21
3.1.4 SUS (System Usability Scale) Test 22
3.2 Feature Extraction 23
3.2.1 Local Binary Pattern (LBP) 23
3.2.2 Histogram of Oriented Gradient (HOG) 24
3.3 Classification 25
3.3.1 Support Vector Machine (SVM) 25
3.3.2 K-Nearest Neighbor Classifier (KNN) 28
3.4 Artificial Neural Network (ANN) 28

3.5 Convolutional Neural Network (CNN) 31
3.6 Cross Validation 33
3.6.1 k-fold Method 33
3.6.2 Leave One Out Method 33
3.6.3 Leave One Person Out Method 33
CHAPTER 4 34
RESULTS OF GENDER PREDICTION 34
4.1 Gender Prediction from Face Images Using Statistical Methods 34
4.1.1 LBP Gender classification Result 35
4.1.2 HOG Gender classification Result 35
4.1.3 Comparison of Feature Extraction and Classification Methods 36
4.1.4 Working on Face Region of the Image 38
4.1.5 Gender Prediction Using Statistical Methods- Conclusion 38
4.2 Gender Prediction from 3D Anthropometric Measurements Using a 3D Camera 39
4.2.1 Data Collection 40
4.2.2 Origin of the Idea 40
4.2.3 Obtained Joint Positions and Features 41
4.2.4 Feature Selection Methods 44
4.2.5 Classification Methods 44
4.2.6 Results and Discussion 46
4.2.7 Gender Prediction from 3D Anthropometric Measurements using a 3D Camera - Conclusion 50
4.3 Gender Prediction from 3D Body Data Using ANN 50
4.3.1 Data Obtained from the 3D Camera 51
4.3.2 Data Expansion 52
4.3.3 Classification Methods 53
4.3.4 Gender Prediction from 3D Body Data Using ANN - Conclusion 60
CHAPTER 5 61
AGE PREDICTION 61
5.1 Age Prediction from Face Images Using Statistical Methods 61
5.1.1 Face Image Database for Age Prediction using Statistical Methods 61
5.1.2 Age Prediction with LBP Feature Extraction Method 63
5.1.3 Age Prediction with the HOG Feature Extraction Method 65
5.1.4. Comparison of Feature Extraction Methods 67
5.1.5 Age Prediction Using Statistical Methods - Conclusion 68
5.2 Age Prediction from Face Images Using Deep Learning 69
5.2.1 Transfer learning with IMDb and WIKI Datasets 70
5.2.2 Transfer learning with Gallagher's Dataset 76
5.2.3 Age Prediction from Face Images - Conclusion 78
5.3 Age Prediction from 3D Body Data 79
5.3.1 Classification Using All-Body Joint Points 80
5.3.2 Classification Using Upper-Body Joint Points 83

5.3.3 Classification Using Lower Body Joint Points 84
5.3.4 Age Prediction from 3D Body Data Using ANN-Conclusion 85
CHAPTER 6 87
A SOFTWARE FOR GENDER AND AGE PREDICTION 87
6.1 Graphical User Interface of the Proposed System 88
6.2 System Usability Scale Test Results 93
CHAPTER 7 97
DISCUSSION AND CONCLUSION 97
CHAPTER 8 101
LIMITATIONS AND FUTURE WORK 101
REFERENCES 103
APPENDIX 115
A. Participants Data of Accuracy Test 115
B. SUS Pie Chart Representations 116

LIST OF TABLES

Table 4. 1. The best percentage of correctness of gender prediction for LBP and HOG feature extraction method and SVM and KNN classifiers and the requirement time to process 30 images 36
Table 4. 2. The accuracy of gender prediction for LBP, HOG and LBP+HOG feature extraction method and SVM and KNN classifiers 37
Table 4. 3. Accuracy of the gender prediction with LBP, HOG and LBP+HOG feature extraction method and SVM classification method are applied only face images 38
Table 4. 4. Accuracies of the gender detection in SVM classification with feature combinations giving the best performance. 48
Table 4. 5. Upper and Lower Joint Points of Skeleton 51
Table 4. 6. Confusion Matrix Using All Body Joints Coordinates 54
Table 4. 7. Confusion Matrix Using Upper Body Joint Coordinates 55
Table 4. 8. Confusion Matrix Using Lower Body Joint Coordinates 55

Table 5. 1. Age Ranges and the distribution of number images with corresponding Gallagher's database. 62
Table 5. 2. Accuracy Rates of LBP with SVM classification method 63
Table 5. 3. Confusion matrix of age prediction with One-versus-one SVM classification method for LBP. 64
Table 5. 4. Accuracy Rates of KNN classification method for LBP. 64
Table 5. 5. Confusion matrix of age prediction with KNN classification method for LBP 65
Table 5. 6. Accuracy rates of HOG feature extraction with SVM classification methods with Linear, RBF and Polynomial kernel functions. 65
Table 5. 7. Accuracy Rates of KNN classification method for HOG 66
Table 5. 8. Confusion matrix of age prediction with KNN classification method for HOG 67

Table 5. 9. The greatest accuracy rates of LBP and HGO feature extraction, SVM and KNN classification methods 67
Table 5. 10. Statistical versus CNN Method on age prediction from Face Images ...79
Table 5. 11. Age ranges, the classes of ANN and the number of people in the age range for age prediction.....80
Table 5. 12. ANN results of All Body Joint Points Features 81
Table 5. 13. Confusion Matrix of LOPO for whole-body 81
Table 5. 14. Confusion Matrix of 10-fold.....82
Table 5. 15. Accuracies of LOPO and 10-fold technique for All Joint Points 82
Table 5. 16. ANN results of Upper Body Joint Points Features 83
Table 5. 17. Confusion Matrix of LOPO for Upper Body Data.....84
Table 5. 18. ANN results of Lower Body Joint Points Features 84
Table 5. 19. Confusion Matrix of LOPO for Lower Body Data 85
Table 5. 20. Best accuracies for Upper, Lower and whole body joint points for all scene data.....86
Table 5. 21. LOPO Results for Lower Body, Joints Upper Body Joints and Whole-Body Joints.....86

Table 6. 1 System Usability Scale 94

LIST OF FIGURES

Figure 3. 1. Procedural Representation of the proposed system 17
Figure 3. 2. Data Collection Step Representation 18
Figure 3. 3. Gender Prediction Procedure.....19
Figure 3. 4. Age Prediction Structure 20
Figure 3. 5. Structure of the System Components.....21
Figure 3. 6. Basic LBP Operator by thresholding [103].....23
Figure 3. 7. Example of LBP based face feature histogram [104] 24
Figure 3. 8. An example of a separable problem in a 2-dimensional space. The support vectors, marked with grey squares, define the margin of largest separation between the two classes. [105].....25
Figure 3. 9. Support faces at the boundary [53] 26
Figure 3. 10. Example of KNN classification [111]. 28
Figure 3. 11. Multiple Input Neuron [113] 29
Figure 3. 12. Three-Layer Network[113].....30
Figure 3. 13. The General Structure of Convolutional Neural Network[130].....31

Figure 4. 1. LBP accuracies for k values of KNN classifier 35
Figure 4. 2. HOG accuracies for k values of KNN classifier.....36
Figure 4. 3. LBP and HOG accuracies for k values of KNN classifier 37
Figure 4. 4.Test Environment in Atilim University.....40
Figure 4. 5. Ideal gender proportions [132] in centimeters: female (left) and male (right). 41
Figure 4. 6. Skeleton Position Joints obtained by Kinect v1 (adopted from [134]). ..42
Figure 4. 7. ANN Structure of the Proposed System.....45
Figure 4. 8. KNN classifier accuracy results for gender detection vs different K values. 47

Figure 4. 9. Combination of features achieving best accuracy with MLP kernel function of SVM.....47
Figure 4. 10. Gender Detection accuracy using ANN with different parameters 49
Figure 4. 11. a) Different Posture b) Different Places where the images captured....52
Figure 4. 12. Two-Layer Network for Gender Prediction 53
Figure 4. 13. Histogram of LOPO method with All Joint Points of Body 56
Figure 4. 14. Histogram of LOPO Method with Upper Joint Points of Body 57
Figure 4. 15. Histogram of LOPO method with Lower Joint Points of Body 58
Figure 4. 16. Histogram of LOPO method with Lower Joint Points of Body With 40 Neuron in Hidden Layer 59
Figure 4. 17. Histogram of LOPO method with Lower Joint Points of Body With 30 and 40 Neurons in Hidden Layer 59

Figure 5. 1. Example images in the Gallagher's database 62
Figure 5. 2. Transfer Learning Representation[145] 69
Figure 5. 3. Training Process with 8 Classes, 50 Images for each class and 20 Epochs 71
Figure 5. 4. Training Process with 8 Classes, 75 Images for each class and 20 Epochs 72
Figure 5. 5. Training Process with 8 Classes, 80 Images for each class and 20 Epoch 73
Figure 5. 6. Training Process with 8 Classes, 80 Images for each class and 40 Epoch 73
Figure 5. 7. Training Process with 8 Classes, 90 Images for each class and 60 Epoch 74
Figure 5. 8. A sample content of the folder where the images of people are labeled at age 2 from WIKI-IMDb Datasets.....75
Figure 5. 9. A sample content of the folder where the images of people are labeled at age 5 from The Images of Group Dataset.....76
Figure 5. 10. Training Process with 7 Classes, all dataset images, 10 Epoch.....77
Figure 5. 11. Training Process with 7 Classes, all dataset images, 30 Epoch.....77

Figure 5. 12. Training Process with 7 Classes, all dataset images, 100 Epoch.....78

Figure 6. 1. The first opened state of the program 88
Figure 6. 2. A person and skeleton lines shown on the camera scene 89
Figure 6. 3. XML structure of the system.....89
Figure 6. 4. Face area coordinate points parts of XML file 90
Figure 6. 5. XML file and calculated heights of a person.....90
Figure 6. 6. Coordinates of some joint points 91
Figure 6. 7. Window with Identification Marks.....92
Figure 6. 8. The Log file text of the system.....92
Figure 6. 9. Gender and age distribution of participants 93

LIST OF ABBREVIATIONS

3D	–	Three Dimension
ANN	–	Artificial Neural Network
CNN	–	Convolutional Neural Network
CPNN	–	Conditional Probability Neural Network
DEX	–	Deep Expectation
DNA	–	Deoxyribonucleic Acid
EEG	–	Electroencephalography
FERET	–	Facial Recognition Technology
FG-NET	–	Face and Gesture Recognition Research Network
FLD	–	Fisher Linear Discriminant
FPLBP	–	Four Patch Local Binary Pattern
FPS	–	Frames Per Second
GMM	–	Gaussian Mixture Model
GPU	–	Graphical Processing Unit
GUI	–	Graphical User Interface
HbE	–	Height by Estimation
HbS	–	Height by Skeleton
HbW	–	Height by Wingspan
HCI	–	Human Computer Interaction
HC-SVR	–	Hybrid Constraint Support Vector Regression
HOG	–	Histogram Oriented Gradient
HR	–	High Resolution
IIS-LLD	–	IIS-Learning from Label Distribution
KNN	–	K Nearest Neighbors
KNN-SVR	–	K nearest neighbors-support vector regression
IFD	–	Indian Face Database
IR	–	Infrared
LBDP	–	Local Block Difference Pattern
LBP	–	Local Binary Pattern
LPP	–	Locality Preserving Projection

LOPO	–	Leave One Person Out
LOO	–	Leave One Out
MAE	–	Mean Absolute Error
MFA	–	Marginal Fisher Analysis
MLP	–	Multilayer perceptron function
NN	–	Neural Networks
ReLU	–	Rectified Linear Units
RBF	–	Radial Basis Function
RBG	–	Red-Green-Blue
RCA	–	Relevant Component Analysis
RGB-D	–	Red -Green-Blue – Dimension
SIFT	–	Scale-Invariant Feature Transform
PS	–	Preprocessing Sequence
SUMS	–	Stanford University Medical Student
SUS	–	System Usability Scale
SVM	–	Support Vector Machine
SVR	–	Support Vector Regression
USA	–	United State of America
WA	–	Washington
VLFD	–	Vital Longevity Face Database
XML	–	Extensible Markup Language

CHAPTER 1

INTRODUCTION

In general, most biometric data are extracted from face images of people [1][2][3]. Nevertheless, some biometrics, such as body shape [4], body mass index [5], height [6], [7], and weight [8], [9] are extracted from body information. Soft biometrics include gender, age, ethnicity, height, and weight, having constituting a widely researched area over the past decade [10]. From soft biometrics, specific information can be gathered to manage resources, better plan services for people and obtain statistical data about the area, such as shopping malls, public areas, sports centers, museums, supermarkets, libraries, etc. Beside gathering such data, security forces also need soft biometrics to identify suspects [11]. Surveillance cameras used to find suspects offer useful information about people; for this reason, images obtained in this way are used to extract data about individuals as well.

The increased number of surveillance cameras installed for safety and control purposes necessitates automatic data acquisition and processing. Otherwise, it is not quite possible to find the person searched for in such huge number of images. Because of the need for automatic data acquisition, the use of computer vision and machine learning approaches come into question. At first, computer vision has an active role in perceiving and tracking people and obtaining their images or body information. Hence, there may be some special hardware or software components of the camera. Then, machine learning tasks including pre-processing, feature extraction and classification take place in this phase and the obtained information passes through certain processes to acquire meaningful data, which may also contain soft biometrics related to individuals.

As the needs, behaviors and appearances of people vary according to age and gender, these two biometric data are generally taken into account. Age and gender of people

can be extracted from both face and body images [12], [13]. Nevertheless, both data acquisition techniques have some advantages and disadvantages. As for face images, resolution is an important factor to predict soft biometrics because low levels of resolution bring about data loss and are too weak to extract features of face. Thus, prediction of soft biometrics from face images need high resolution images. On the other hand, with high-resolution face images, age and gender prediction is enviable as also supported by the study [14]. For body images, soft biometrics are predicted using anthropometric measurements [15],[16], gait-base [17][18] [19] or gesture-based [20] measurements. In addition to standard RGB cameras and images, 3D cameras and RGB-D images with deep data can help to extract soft biometrics. However, the problem with this technique is the lack of standard datasets and researchers studying body information to predict age and gender have to create their own dataset.

Automatic age range and gender prediction is an important issue to identify people in computer-based systems. These are two special areas in soft biomatrices, with an area of application encompassing security systems, human computer interaction (HCI), law enforcement [21], etc.

Prediction of age and gender of people can be done either through an appearance-based approach (face images, body images, body metrics, clothing, etc.) or a nonappearance-based approach (handwriting, social media etc.). In this thesis, two of the appearance-based prediction approaches; face images and body metrics are examined.

1.1 Motivation

As stated previously, predicting the age and gender from face images and the body information has been a subject of research by many experts. However, it is still an unsolved problem and should be considered with different conditions. One of the conditions is that the face is not seen while full body data is captured. For standard RBG cameras, face images can provide features to predict age and gender. However, when the face is not seen, anthropometric measurements or postures may be used to predict age and gender. In some cases, people purposefully hide their faces in the presence of cameras in order not to be identified.

Another problem occurs in crowded areas, where whole-body images cannot be captured. In many studies, gait-based age and gender prediction has been investigated,

requiring at least a period of one step walk [22][23], whose caption cannot be always accomplished in crowded areas.

Similarly, the anthropometric measurement-based technique requires capturing the whole-body data for the purpose of prediction. For this reason, better and more effective methods are needed to make predictions using half of the body information or when the face image or whole-body information is not present.

Last but not least, predicting age and gender from face and body information may have been integrated in some programs, but there is no directly available program for this purpose only. To overcome these obstacles, the present thesis is an attempt to develop and test a software tool that can make age and gender predictions from facial images, whole body data or part of the body data.

1.2 Research Problem

Age and gender prediction can provide useful statistical information with areas of use in a variety of forms such as security [24], surveillance [25], health [26], human computer interaction [27], social networks [28], to mention a few. For age estimation, learning algorithms cannot expect "sufficient and complete training" data because aging is a smooth and slow process, and close ages look quite similar [29].

To extract the necessary information for age and gender prediction, the present thesis employs two types of cameras; standard RGB and 3D cameras. For the biometric trait and features of faces, standard RGB cameras are needed. However, standard RBG cameras can overlook faces and bodies due to occlusion effects. Instead, a 3D camera is used and tested experimentally so as to determine to what extent it affects the accuracy of age and gender prediction. 3D cameras can capture video sequences and provide in-depth information about bodily joint points. As for the present thesis, we propose the use of both cameras to predict age and gender.

In order to develop the system, both the latest and traditional machine learning techniques are applied and compared in terms of accuracy. For gender prediction, the traditional statistical methods, LBP (Local Binary Pattern) and HOG (Histogram Oriented Gradient), are applied on face images for feature extraction and extracted features are classified with KNN (K-Nearest Neighbor) and SVM (Support Vector

Machine). Additionally, 3D biometric data obtained from 3D camera is used for anthropometric measurements with statistical SVM and KNN classifiers, and only the posture information of the body is classified with ANN. For age prediction, both the statistical LBP and HOG methods are applied on face images and classified with KNN and SVM, while convolutional neural network method is used to predict age from face images. Moreover, 3D body information is applied to predict age range. In the end, a software tool is generated combining both age and gender prediction features.

The present thesis offers the following research question:

- Can the combination of two different sets of biometric information be used to predict age and gender of people?
- Which of the CNN or statistical approaches is better than the other?
- Can 3D body information be used for age and gender predictions? What is the accuracy of the proposed solution using such information?
- Can only the whole-body information be used for age and gender prediction or should the information regarding upper and lower body parts be also considered in the classification phase?

To sum up, this study investigates the applicability of the combination of face images and 3D body information for automatic age and gender prediction of individuals.

1.3 Outline

The present section explains the topic, motivation and the problem statement as to age and gender prediction using the data obtained from RGB and 3D camera. Chapter 2 offers the background information about gender and age prediction from face and body, focusing on studies related to specific techniques developed in the literature so far. In Chapter 3, first the core structure of the study is explained, followed by feature extraction methods (LBP and HOG), the classification methods (KNN and SVM), Artificial and Convolutional Neural Network (CNN) their properties and structures. Later, the results of the gender prediction from face images and body information are given in Chapter 4 by dividing gender prediction from the body information into two sections to manage the results according to the statistical and neural network methods. For age prediction, again the face images and body information are examined separately. The results are presented in Chapter 5, with the combination of gender and age prediction tools in the form of one software explained in Chapter 6 to also include

usability tests results. In Chapter 7, the conclusion and discussion of the findings are given. Finally, the future work and limitations are itemized in Chapter 8.

CHAPTER 2

LITERATURE REVIEW

In recent years, automated gender and age prediction have been among frequently used applications of security systems, human computer interaction (HCI), law enforcement [21], etc. There are different methods to predict gender and age of humans, some of which are stated in the following sections. In this study, gender and age prediction are considered according to the facial and body metric information of individuals. Initially, we will present some studies from the literature about gender prediction using face image and, then gender prediction using body metrics. Next, certain studies are mentioned about age prediction using face images.

2.1 Gender Prediction

Automatic gender prediction is the process of determining the gender of a human according to the characteristic properties that represent his/her masculinity or femininity. Gender detection is used in many areas such as personalization and recommender systems [30], behavior analysis [31], consumer research [32], digital forensics [33], security and biometrics [34], human computer interaction [35], mobile applications [36], etc.

Traditional approaches for gender recognition include face [35], and handwriting analysis [33]. Some studies have also shown that the body posture, too, can be used for gender classification [17], [18],[37]. The approaches for gender detection based on data derived from human body can be classified as either appearance-based and nonappearance-based [38]; the former uses static body features (face, fingernail, body shape [39], etc.), dynamic body features (gesture, motion, gait etc.), and apparel features (clothing, footwear, etc.); whereas the latter employs biometric features (fingerprint, iris, ear, skin color [40], etc.), bio-signals (DNA, EEG, etc.), and social

information (blog, email, handwriting, etc.). The most commonly used features for gender detection are the features gathered from the face [41], [42]. which is used not only for gender detection, but also for subject identification, age detection, emotion prediction, etc. In order to use face images for gender detection, high resolution (HR) images are needed from the right angle that capture all parts of the face. However, in daily life it is not always practical to obtain HR face images because of the distance between the face and the camera as well as the orientation angle between the camera and the subject imaged. Finally, when some parts of the face are occluded, the face image might not be necessarily usable for gender detection. In order to overcome this difficulty, whole-body properties are used for gender detection. While considering these properties as input for gender detection, different measurements such as anthropometric [43], gait and motion properties [7], and kinematical, dynamical, and motor control parameters of skeleton and limbs [44] can also be used. Low-cost range sensors such as the Kinect sensor (Microsoft, Redmond, WA, USA) are able to track the orientation and position of a human body and its limbs, and can be used to measure the anthropometric dimensions of an individual [45]. The accuracy of measurements performed by the Kinect sensor has also been validated with metrological tools [46]. Since face images and 3D body images are used for gender prediction in this thesis, studies using those data are summarized in the following subsections.

2.1.1 Gender Prediction from Face Images

Gender prediction from face images is among appearance-based approaches. Because the face is a static feature and face images can be obtained from a long distance, the most usable biometric feature is the face. There are numerous areas of application for gender prediction using face images [47]. Different methods have started to be used with the understanding that machines can predict gender using face images.

In 1990, Golomb et al. [48] proposed a gender prediction method in order to determine the gender of a human by machines. The proposed method is a neural network approach with 90 compressed images (45 males and 45 females). The average error rate is calculated as 8.1% with an error of 10 test images arranged in a fixed size 30x30. This study is important because it is among the first works accomplished in the field of gender prediction by machines. Gender prediction was first performed using the neural network, continued with a four-layer Convolutional Neural Network (CNN) with the study [49], and the obtained accuracies are 98.75% for the Stanford University

Medical Student (SUMS) database and 99.38% for AT&T database. The main outcome of the studies is that CNN is used for gender prediction in real-time application, but it is computationally expensive.

The graph of face with a two-dimensional view [50], like geometric features face edges labeled with distance vectors [51], hybrid classifier with Radial Basis Function (RBF) network and decision tree [52] are proposed to predict gender during a 10 year period. Then, the Support Vector Machine (SVM) approach is used for gender prediction in 2000 by Moghaddan and Yang [53]. The study shows the results of SVM with kernel functions; RBF, Fisher Linear Discriminant (FLD), Linear and Quadratic classifiers to predict the gender of face images. The lowest error rate belongs to the Gaussian RBF kernel function with an overall error rate of 3.38% on the dataset, which has 1800 images at a resolution of 21x21.

In 2005, Jain et al. [54] proposed a method with Independent Component Analysis as a feature vector of image, and SVM as a classifier. 500 (250 males and 250 females) images are used from the FERET [55] dataset, with 96% accuracy. After this study, there have been many gender prediction methods using the SVM classifier [56] which apply Local Block Difference Pattern (LBDP) and obtain effective results [57].

In 2006, gender prediction with LBP feature extraction method is used by Sun et al. [58], Lian and Lu [59] and Erno and Loope [47] with SVM, Self-Organizing Map (SOM) and Adaboost classifier methods on the FERET database, with accuracy rates at 95.75%, 96.75% and 91.11%, respectively.

In 2010, Nazir et al. [60] proposed a gender prediction method which crops the face area from the images by Viola and Jones face detection method [61]. To normalize the illumination, histogram equalization is applied. The Discreate Cosine Transform (DCT) feature extraction method and KNN (K-Nearest Neighbor) classification method are used to predict gender on the SUMS image database. The highest accuracy obtained is 99.3%.

In 2013, Shih [62] proposed a feature extraction method, known as Precise Patch Histogram (PPH), to increase the accuracy of gender prediction, with a performance accuracy of 94.8% on Labeled Faces in the Wild (LFW) dataset [63].

In 2014, Liu et al. [64] proposed gender prediction method using LBP and Histogram-Oriented Gradient (HOG) feature extraction methods with SVM classifier. For LBP, 91.43%, for HOG 94.38% and for LBP+HOG 94.88% accuracies were obtained with the SVM classifier on LFW face database.

In 2017, Khalifa and Şengül [65] proposed LBP and HOG feature extraction methods and SVM and KNN classification methods to predict gender. The most accurate result is 98.79% with the SVM classifier using a combination of LBP and HOG feature extraction methods on FERET and University of Texas at Dallas (UTD) databases. Their method has been the latest and most precise techniques of predicting gender.
These studies show that gender prediction can be done with high accuracy using face images. Therefore, in this study, face images are examined to predict gender.

2.1.2 Gender Prediction from Body Information

Body-based gender detection studies can be grouped into four categories: 1) gait based, 2) anthropometric measurements-based, 3) motion-based, and 4) combined approach using gait and anthropometric measurements together. In this section, we first and briefly examine the gait-based and gesture-based approaches and, then, focus on the anthropometric measurement-based approaches as well as combined studies using gait and anthropometric measurements.
Body motions during walking have been analyzed to predict gender by many studies with image-based gait analysis. Among them, [66], [67] and [68] achieved the accuracy rates of 83%, 94% and 96%, respectively. High accuracy rates are achieved even without using HR images. Subject identification and gender detection can be done with the help of human silhouettes while walking [67].
For gender detection, Cao et al. [17], Guo et al. [18] and Collins et al. [19] used all the body parts identified from 2D video images to extract the walking features of the person and classify the data according to the gender. Also, Miyamoto and Aoki [69] used the time series variation, which was normalized with linear interpolation in the joint positions taken from the Kinect v2, to determine significant features for gender detection. In the study, they applied not only the three-dimensional coordinates of joint positions, but also the projections of joints from a 3D space onto the 2D plane according to the exposure angle. The authors, in addition, used SVM as the classifier to predict the gender, thus reaching the accuracy of 99.12% with 12 subjects (six males and six females). Although gait-based approaches do not require HR images, one of their drawbacks is that they require the images of a full walking period of the person, which is not always possible in crowded urban environments.

To predict gender, gestures have been used by Won et al. [20]. In the study, nonverbal behaviors were also emphasized to do the prediction. With different postures of twelve men and twelve women, gestures were captured in front of the Kinect camera and used as training data. The angle between the shoulders and the neck of human gestures were extracted as movement features, and the length of the body parts were extracted as static length features for classification in ten different postures. After using the standard technique to reduce the number of features, the top ten features were selected to classify gender, reaching an accuracy of 83%.

The anthropometric measurements-based gender detection approaches can be divided into 2 groups. The first group uses 2D anthropometric measurements, while the second group uses 3D measurements. The first study of anthropometric measurements in 2D was proposed by Adjeroh et al. [70] with the analysis of the correlation of anthropometric features from the CAESAR database [71]. Also, using anthropometric features, the copula model was proposed using the same dataset by Cao et al. [16] for gender and weight prediction. The proposed method uses both the copula model and SVM in combination, and the accuracies are 99.1% for only body feature, 87.4% for only head feature, and 99.4% for both head and body features.

To predict gender from still images, human metrology has also been studied with ratios of anthropometric measurements for the LUPI (Learning Using Privileged Information) framework by Kakadiaris et al. [15], followed by a similar approach to Cao et al. [16]. In [16], the ratios of the anthropometric measurements are more accurate than the actual measurements, allowing an accuracy of 98% for the LUPI framework. The study also shows the result of 3D pose estimation algorithm to obtain joint locations in real images from the PaSC and SARC3D datasets from 2D annotation. The best accuracy value for real images was 86%. However, the depth information was not used to measure the anthropometric measurements.

There are only a few studies using the 3D anthropometric measurements one of which has been done by Sandygulova et al. [72], where children's gender and age prediction was experimentally tested. A total of 276 boys and 152 girls between 5 and 18 years were involved in the experiments. Standard machine learning approaches were applied on features, which are obtained with Kinect by modelling 3D body metrics. The system performance was evaluated on volunteers using the adaptive social robot in wild. For gender detection, shoulder height and hip to shoulder proportions were important

features for classification. The gender detection accuracy was 73% in a real-time adaptive robotic system.

Another gender detection approach using the 3D anthropometric measurements has been suggested by Andersson et al. [73], where gender prediction and body mass index prediction were investigated using anthropometric and gait information together. Eighty features were used for classification: the Euclidean distance of tracked joints (20 attributes), and gait attributes such as spatiotemporal (4 attributes) and kinematic (56 attributes). For training and testing, images captured from 44 people were used for each set. For the classification part, KNN (K=3), SVM and multilayer perceptron (MLP) were used. The most accurate result (95%) was obtained by MLP using all attributes. Also, using the anthropometric measurements only, MLP achieved an accuracy of 93 %.

According to the studies stated in this section, not only the face images, but also the body image and body information provides data to predict gender. For this reason, in this study, joints point coordinates of the body is used to predict gender from person's body.

2.2 Age Prediction

Automatic age prediction is the process of classifying the age of the people from some personal characteristics such as text messaging patterns [74], social media [75][76], online behavior [77], and face images [78]. For age prediction, face image is the widely used characteristic. As in many studies, in present thesis study, face images are used to predict a person's age. Biometric features of a face change day by day. Besides age can be determined with the help of bone movements, growth and skin-related deformation with wrinkles and reduction of muscle strength [79]. The age- related features of a face help to estimate the age from the face. However, humans cannot accurately estimate the age because of external factors, such as health, living style, make-ups and environment, and weather conditions [80]. Because estimating age also changes according to genes and external factors, it is still a challenging problem for computer vision.

Still, there are some applications to help with estimating age or age ranges. As Han et al. [21] mentions, law enforcement, security control and Human Computer Interaction (HCI) are some of the potential fields of application for automatic age estimation. As

in law enforcement, searching for suspects according to the estimated age from a police record archive is one of the applications of automatic age prediction. For security control, people who are younger than 18 can be prevented from purchasing alcohol and cigarettes and accessing inappropriate web pages. As HCI, the system content or the advertisements can be arranged according to the users' age. Moreover, there are also other reasons to use automatic age estimation, as Lanitis [81] states, such as in data mining and organization, classifying, storing and retrieving age-based face images from e-photo albums and the Internet.

Chao et al. [82] in 2013 proposed a new age estimation approach with three contributions to estimate an individual's real age. First, to explore the connections between facial features and age labels, they combine distance metric learning and dimensionality reduction. Then, in order to establish an intrinsic ordinal relationship among human ages and to cope with the potential data imbalance problem, a label-sensitive concept and several imbalance treatments are introduced into the system at the training phase. Finally, to capture the complicated facial aging process for age determination, an age-oriented local regression is presented. To reduce the dimensionality of the features and preserve the most important information for age estimation, locality preserving projection (LPP) is utilized. A label-sensitivity concept, which importance the label similarity during the training phase of LPP, is introduced to exploit the ordinal relationship among age labels. KNN-SVR (K nearest neighbors-support vector regression), an age-based local regression algorithm, is presented to capture the complicated facial aging process. The proposed method was examined in several experimental settings on the FG-NET [83] aging database. Moreover, other than LPP, the label-sensitive concept on dimensionality reduction algorithm is also applied on a popular algorithm, called the "marginal fisher analysis" (or MFA). For distance metric adjustment and dimensionality reduction, the relevant component analysis (RCA) and LPP are used, which are both learning machine algorithms requiring a training phase. From the simulated results performed on the FG-NET database with different experimental settings, the newly proposed method has the lowest MAE comparison to the state-of-art algorithms.

In 2013, Liu et al [84] attempted to improve age estimation with a novel viewpoint, taking advantage of both the supervised training data and human annotations. For this

purpose, first, a fuzzy age label is defined and the traditional data labels are merged with it into the Support Vector Regression (SVR) framework. In the deterministic label, traditional age label has a fixed age value. However, in the fuzzy label, the lower and upper bounds of the ground-truth age are annotated, after which the next step is to solve hybrid labeled learning. Then, the problem is re-formulated with the standard SVR for regression-based facial age estimation to be solved with the existing SVM solver, such as LIBSVM with minimal changes. In the study, the FG-NET aging dataset and 5 volunteers' face images are used to label their age and, then, acquire the fuzzy labels. Nine (9) key points, 4608-d SIFT features, and Gabor features are computed and reduced into 120-d vectors by PCA, and 1152-d GMM (Gaussian Mixture Model) features are trained. The proposed Hybrid Constraint Support Vector Regression (HC-SVR) method is compared with standard SVR, AGES and LARR age estimation algorithms, and the results reveal that minimum Mean Absolute Error (MAE) is 5.28±0.05 and belongs to HC-SVR. Consequently, the proposed method has the best overall performance among all the three features.

Again in 2013, Geng et al. [29] proposed an instance associated with a label distribution which covers a certain number of class labels, each of which describes the instance degree. In this way, a face image can exploit learning both chronological age and adjacent ages. For this purpose, IIS-Learning from Label Distribution (IIS-LLD) and conditional probability neural network (CPNN) algorithms are proposed to learn from such label distribution. The first label distribution algorithm is IIS-LLD, which is the maximum entropy model, and the second one is conditional probability neural network (CPNN) which models the entropy by a three-layer neural network. The proposed CPNN is similar to the Modha's neural network [85] an unsupervised learning algorithm for conditional Probability Density Function (PDF). However, CPNN is trained in a supervised manner. After applying the learning algorithm, the authors ranked all class labels according to their description degrees to estimate the exact age, and the maximum description degree is the predicted age. The maximum sum of description degrees of all the ages within an age range may be the predicted age range as stated in the study. In the experiments, two datasets were used, namely FG-NET and MORPH [86], with 55,132 face images from more than 13,000 subjects range from 16 to 77 with a median age of 33. According to the findings the advantages of label distribution algorithm are as follows: first, the prior label distribution of the

training samples contribute to the learning of multiple classes; second, the algorithms tend to learn the similarity among the neighboring ages without considering the prior label distribution.

In 2009, Gallagher and Chen [87] proposed an age and gender image database which are obtained from in a group of people. The proposed system not only considers features of people's images, but also contextual features of their facial structures to determine demographic information such as age and gender. The idea of the proposed method is that the spatial position of the group of people provides information as to the social context inside the images. For the face parts of the images, contextual features are used with absolute position, relative features, minimum spanning tree and nearest neighbor methods. A combination of contextual and appearance features is used and the most accurate result is 42.9% for age, and 74.1% for gender prediction.

Fazl-Ersi and Awad [88] proposed an age and gender recognition framework using LBP, Scale-Invariant Feature Transform (SIFT), Color Histogram (CH) feature extraction methods and SVM classification methods with the RBF kernel function. The Gallagher's data set [87] is used to train and test the system with different feature extraction methods to compare and improve the accuracy of age and gender recognition. Three feature vectors that are generated using LBP, SIFT and CH are concatenated and gathered into a single feature vector, with only 200 informative features selected. The results show that the combination of three different feature extraction methods gives the best result with 91.59 % for gender and 63.01% for age accuracy. The age ranges are divided into five classes, which is different from Gallagher's proposal.

Eidinger et al. [89] proposed an age and gender estimation benchmark tested in Gallagher's database. The proposed study first detects the face, then aligns it according to the reference coordinate frame and Four Patch LBP (FPLBP) [90], which is an LBP based feature extraction method applied to classify age and gender. In order to avoid overfitting, the dropout-SVM [91] method is used. In the study, LBP and FPLBP feature extraction methods individually and in combination are tested with the SVM classification method. The results show that the combination of LBP and FPLBP with the dropout-SVM gives the best accuracy at 66.6% for age classification and 88.6% for gender classification on Gallagher's database.

Azarmehr et al. [92] also proposed an age and gender classification framework, which is real-time embedded and based on videos. As similar to Eidinger's method, first the face area is detected in the image and, then, the face part is cropped and aligned using the nose and the eyes. Then, the illumination effect is reduced using the Preprocessing Sequence (PS) approach [93]. After the pre-processing steps, the LBP feature extraction method is applied by dividing the image into non-overlapping regions. Extracting the LBP of each cell region histogram, known also as LBP Histogram or LBPH, the feature vectors are concatenated and a single-feature vector is obtained. In order to train the system, the FERET, MORPH and Gallagher databases are used and for testing, the FERET [94], Adience [89], BioID [95], and PAL [96] databases are utilized. For four age ranges, age accuracy is 80.7% with dropout-SVM classification method with RBF kernel function on Adience database.

In 2016, Khalifa [14] proposed gender and age prediction with LBP and HOG feature extraction methods using SVM and KNN classification methods. The proposed techniques are the best in overall performance of age and gender prediction from face images at 99.87% and 100%, respectively, on three databases widely used for this purpose: FERET, FG-NET and UTD. In general, first the preprocessing steps are applied on face images, then the feature extraction and concatenation of feature vectors are performed. Finally, the classification process is executed and accuracy rates are calculated.

Because the previous studies have high accuracies in age prediction using face images, it is also examined in our study. Although face images accuracy for age prediction is high, body information is also used to predict age in this study.

CHAPTER 3

MATERIALS AND METHODOS

Nowadays, it is becoming increasingly common for people to use and work with computers or embedded systems in many areas. The most fundamental reason for this is that these tools have the ability to make decisions or perform operations independently of the person, having acquired human-independent working abilities by utilizing "learning" features. In other words, learning and making decisions, which was once considered as specific characteristic of humans, has now become an ability possessed by machines as well. In fact, what is being carried out in the new world of technology is machine learning, having become a science in itself and enabling machines to make decisions for a specific purpose through statistical methods.

In this respect, two basic methods are used to improve the learning ability, supervised and unsupervised learning. The former is similar to teaching a student to find the solution to the next similar problem, by showing the result of the previous problems. In other words, supervised learning is the realization of learning by increasing the accuracy of the training phase through giving input and expected outputs. While supervised learning leads to learning by giving inputs and outputs, only inputs are given in unsupervised learning. Because unsupervised learning is performed by clustering inputs without using outputs, there is no training phase and, hence, it has a more complex structure than supervised learning. If there are problems that can be reached on output, suggestion is to use supervised methods. However, unsupervised learning techniques should be preferred for problems without outputs.

Since in the present study, we have labeled learning data for age and gender estimation problems, so supervised learning methods are preferred for classification purposes, while both statistical methods and neural network methods are applied to classify age and gender of humans.

3.1 Proposed Gender and Age Prediction System

There are numerous applications for gender and age prediction of humans; these include human computer interaction, resource management, security systems, statistical data analysis, and others. To automate the prediction of age and gender of people, there are many studies - as stated in the previous sections - but there is no software proposed for this purpose as yet. In order to create a software which can automatically determine the gender and age and provide statistical data about people who pass in front of the camera, we have proposed a system whose procedural representation is shown in Figure 3.1.

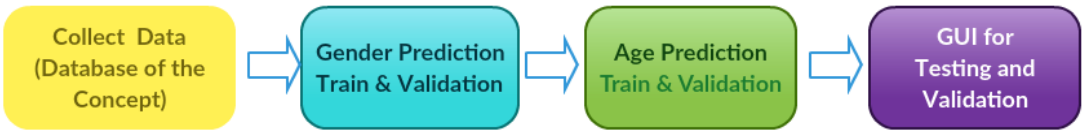


Figure 3. 1. Procedural Representation of the proposed system

There are four main parts in the study: data collection, gender prediction, age prediction, and GUI (Graphical User Interface) for the validation and testing of the system with SUS (System Usability Scale) test to understand the usability of the system. In what follows, each part of the system is explained in detail.

3.1.1 Data Collection

For problems which depend on classification-based supervised learning, the data is important for the training phase of the classification. If the amount of data in the training phase is large and clear, in general the accuracy rate of prediction also becomes high. For this purpose, databases that are collected and used in previous studies are useful and provide comparable results. However, it may be difficult to find data sets in each field. In such cases, collecting data is an important issue.

In this study, the face images and 3D body information are used to predict age and gender. The data collection phase for this purpose can be explained as in Figure 3.2

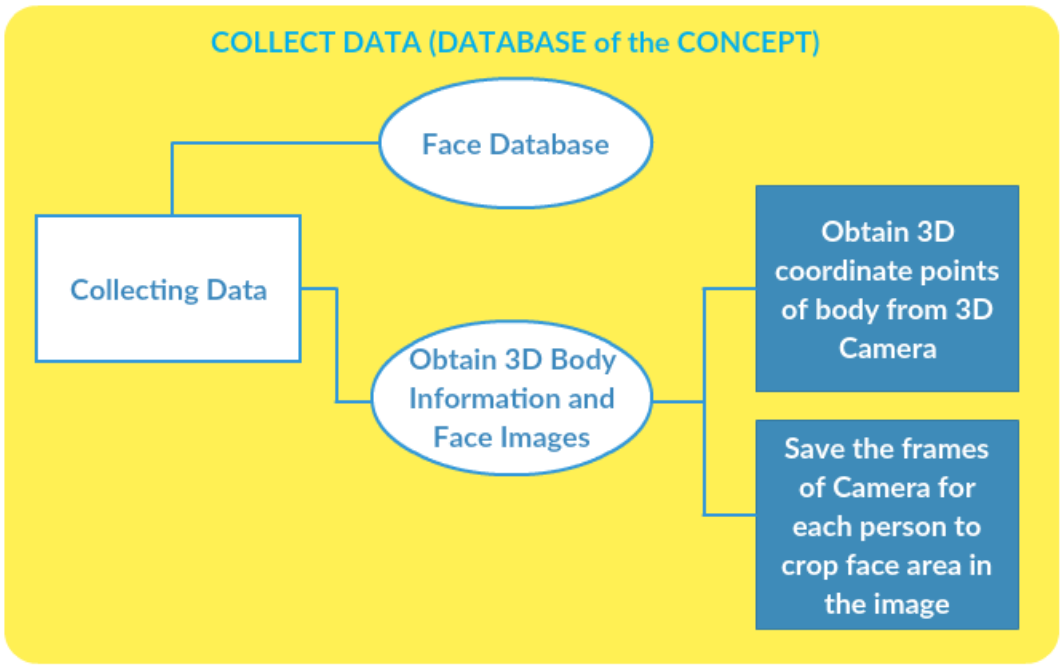


Figure 3. 2. Data Collection Steps Representation

For classification based on face images, there are a large number of databases which supply labeled gender and age images of faces. For gender prediction from face images, the FERET database is used to train the classification methods. Nevertheless, for gender prediction from 3D coordinates of joint points, there is no publicly available database, so we collected our training data. Therefore, using the 3D camera, the 3D coordinates of the joint points of body were collected from 170 people between ages 16 and 75, 65 female and 105 males. Additionally, each frame obtained from the 3D camera has been saved to test the classification based on face images by means of later comparison or else.

To collect the information from individuals about their appearance and body joint points, an ethical committee report was published by Atilim University, finding it appropriate to collect such data. However, it was difficult to collect data from a target group less than 18 years old because both parents should give consent. The next step is age prediction, and the body database is prepared though it lacks a satisfactory number of persons under 18. First, age classification is studied on the face images in the system. The images of Groups Dataset proposed by Gallagher et al. [87] are used in SVM and KNN classification methods, and then in CNN classification method for age prediction from face images. The data collection phases and conditions appear in the following sections in more detail.

3.1.2 Gender Prediction

There are many gender prediction methods based on different data obtained from humans. In this study, our focus is gender prediction from face images and 3D body joint points data. For this purpose, the procedures can be defined as shown in Figure 3.3.

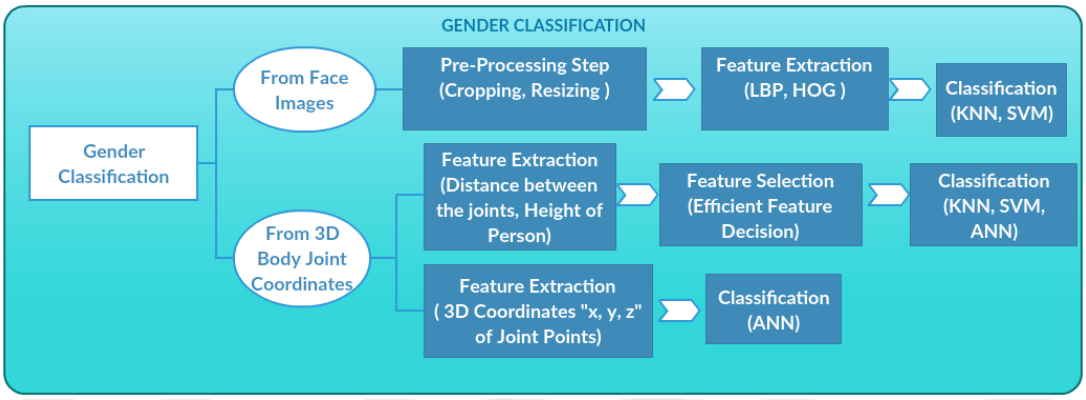


Figure 3. 3. Gender Prediction Procedure

One of the approaches used in this study is gender prediction from face images based on statistical classification methods. Firstly, pre-processing steps such as cropping the face from the image and resizing are applied on the face images. Later, the feature extraction methods, as stated in Section 3.2, Local Binary Pattern (LBP), and Histogram Oriented Gradient (HOG) are used on each face image after pre-processing. The final stage is classification, performed with extracted features. The K-Nearest Neighbor (KNN) and Support Vector Machine (SVM classification methods) are explained in detail in Section 3.3. These are the statistical methods which have also been applied for the age prediction methods.

In this study, gender prediction is not only considered based on the face images, but also on the 3D coordinates of the body joint points. In the part introducing these coordinates, both statistical methods: SVM, KNN and an Artificial Neural Network (ANN) are used for classification. There are two approaches using the body data: one of them is finding the distance between two joint points and the other one is directly using the joint points themselves. For the first one, the distances between two joint points and three differently calculated heights have all been used as features, totally 193, and using all the features together does not affect the classification phase positively. Therefore, feature selection is applied to obtain more accurate results in the classification process. A detailed explanation of feature extraction and selection methods comes in the following sections.

The second approach for using features is the 3D coordinates of joint points for ANN. Because ANN selects features in the training and validation phase, directly using the joint point coordinates improves the accuracy rates, in which case, the upper and lower parts of the body joint points are tested separately with ANN classification. Section 4.3 contains the results of each approach.

3.1.3 Age Prediction

Another main component of the system is age prediction, and the structure of studying this element appears in Figure 3.4.

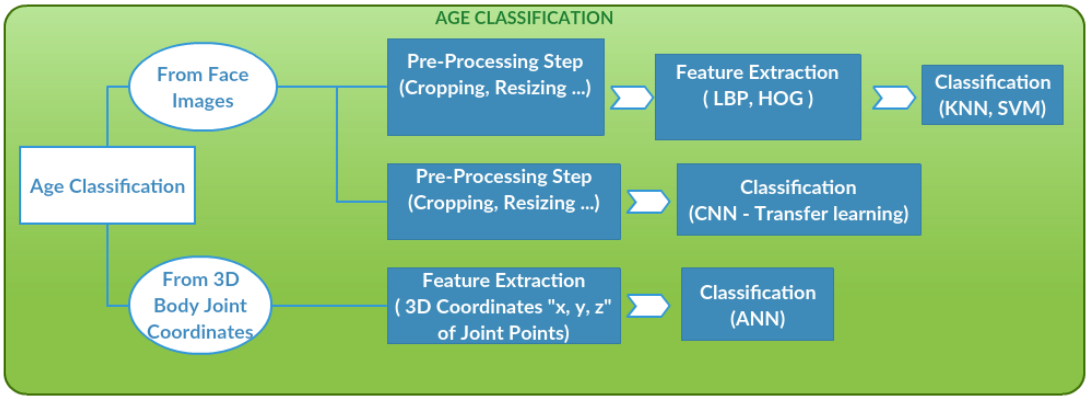


Figure 3. 4. Age Prediction Procedure

The present work aims to predict the age of the people according to the face and body data. For age prediction from face images, two approaches have been considered for classification methods; statistical methods (SVM and KNN using LBP and HOG feature extraction methods) and Convolutional Neural Network (CNN). Both of these methods need to have pre-processing steps to crop the face area from the images and resize the area in order to yield an equal number of features for feature extraction.

For age prediction from the body, 3D coordinates of joint points are extracted for ANN. Because ANN selects features in the training and validation phase, directly using the joint point coordinates improves the accuracy rates. Under these conditions, the upper and lower parts of the body joint points are tested separately with ANN classification.

All experimental results are shown in Chapter 5 in detail.

3.1.4 Graphical User Interface

In our system generated for testing the validation and usability scale of the proposed methods, there are two major and four minor functionalities. Figure 3.5 shows the structure of the components in the system.

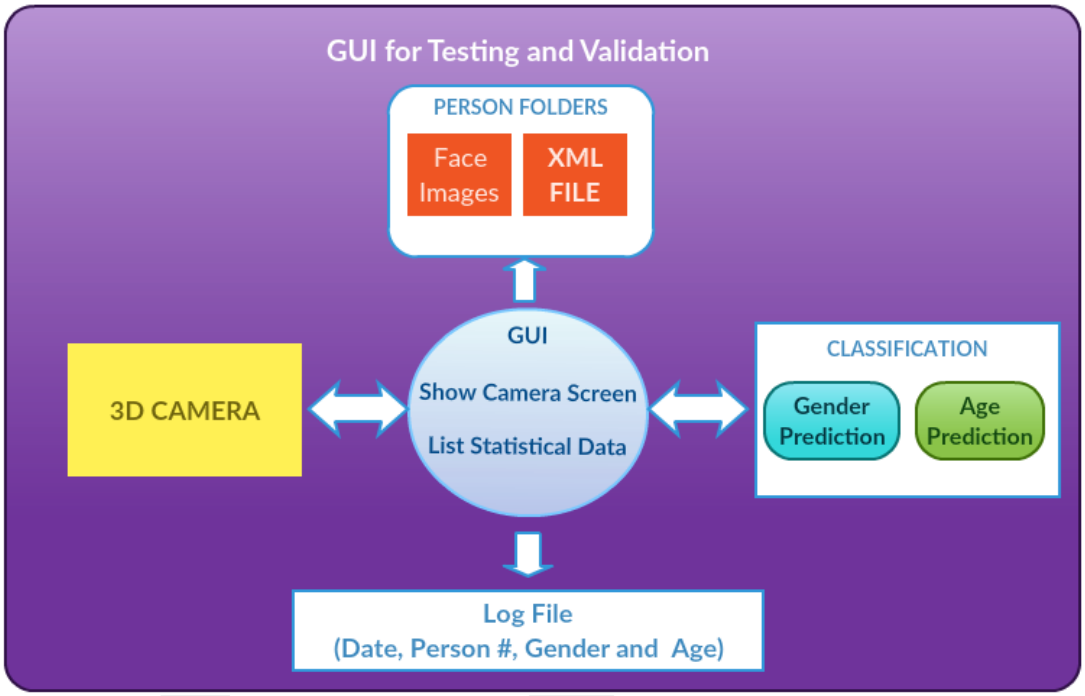


Figure 3. 5. Structure of the System Components

The major functionalities cover gender and age prediction according to the data captured from the 3D camera. For gender prediction, the 3D joint coordinate points of individuals are tested using pre-trained ANN. As for age prediction, a face-based method is used by cropping the face area from the whole image captured from the RGB camera; later, resized face images are tested using pre-trained CNN. The minor functions of the system include listing the people captured by camera, saving a log file of these individuals, presenting the statistical age and gender data captured by camera and showing the last and the selected persons data. A detailed and illustrative description of the user interface appears in Section 6.1.

3.1.4 SUS (System Usability Scale) Test

This was initially proposed by Brooke in 1996 [97] in order to scale system usability at a global or general level. Up to that time, each usability metric was content-specific, but a global usability metric was presented through the Brooke's method for the first time. The scale has played a key role in providing a common evaluation of many systems and comparing system usability. There is a ten-item scale defined as Likert-type with 5 point scales from "strongly agree" to "strongly disagree", as per Table 3.1.

Table 3. 1 System Usability Scale

	1	2	3	4	5
1. I think that I would like to use this system frequently.					
2. I found the system unnecessarily complex.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system to be well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

As can be seen in the table, the statements have been generated in terms of usability and scaled bearing in mind different aspects such as training, support and complexity. Thanks to this scale, one has the opportunity to evaluate the system's usability from a broad perspective as in many other studies, such as the Internet of things applications [98] and mobile applications [99]. For this reason, the proposed gender and age prediction system is tested using SUS by 80 volunteer senior students in Computer engineering, Software engineering and Information Systems engineering departments at Atilim University, Ankara. The results of the scale are analyzed in Section 6.2.

3.2 Feature Extraction

Feature extraction [100] is a method used before classification in order to reduce redundant data and obtain more useful features from hidden information as well as to reduce the amount of data processed due to the dimensionality problem. The following feature extraction methods are mostly used in the face features.

3.2.1 Local Binary Pattern (LBP)

In gender prediction, one frequently used texture-based feature extraction method is the Local Binary Pattern (LBP) method where, rather than the global information, comparisons of local pixel values are used to gather better performance on gender classifications [101]. The method was originally introduced by Ojala et al. [102], who coded the local structure around each pixel. In 3X3 neighborhood pixels, the centered pixel is compared with each one of its adjacent pixels, if the centered pixel is less than or equal to the pixels it is being compared with, the new value for the current pixel will be "1"; otherwise, "0". The new binary values are concatenated in a clockwise direction.

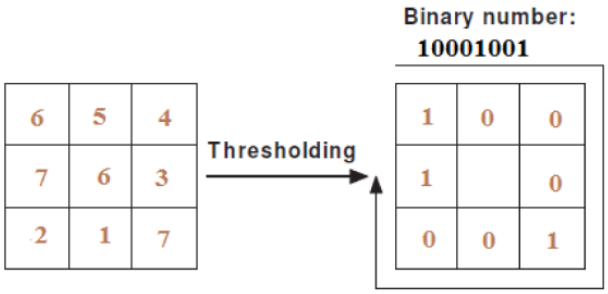


Figure 3. 6. Basic LBP Operator by thresholding [103]

In Figure 3.6, centered pixel "6" is compared with its adjacent pixels, and the binary values of the adjacent pixels are shown on the related position. The new binary number "10001001" is generated after the LBP operation. The value of the LBP code for the pixel $(\boldsymbol{x_c}, \boldsymbol{y_c})$ is defined by;

$$\boldsymbol{LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{P-1} s(i_n - i_c)2^P} \tag{3.2}$$

where $\boldsymbol{i_c}$ is the gray-level value of the centered pixel, and $\boldsymbol{i_n}$ is the gray-level values of the P surrounding pixels *n=0, ..., P-1*, with radius R. The function $\boldsymbol{s(x)}$ is also defined as:

$$s(x) = \begin{cases} 1, & if\ i_n \geq i_c \\ 0, & if\ i_n < i_c \end{cases} \tag{3.3}$$

$LBP_{P,R}$ is described as the LBP operator with a neighborhood size of P equally spaced pixels on a circle of radius R. For the face images, small-size patterns occur; hence, it is useful to apply LBP histograms. To extract the LBP histograms, face images are divided into local regions. Then, all LBP histograms are concatenated and the feature vector is obtained as shown in Figure 3.7.

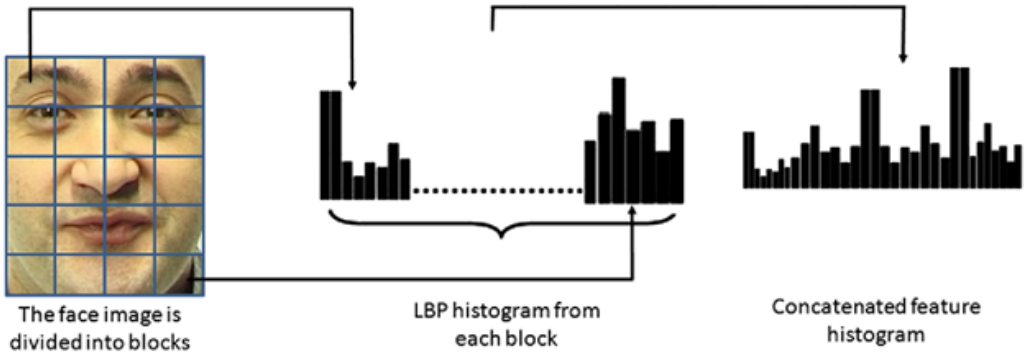


Figure 3. 7. Example of LBP-based face feature histogram [104]

3.2.2 Histogram of Oriented Gradient (HOG)

The Histogram of Oriented Gradient (HOG) is another feature extraction method which uses texture feature by counting the occurrences of gradient orientation in localized portions of an image. To extract these features, vertical and horizontal gradient values are used to calculate magnitude m and orientation θ for each pixel in the image using the following formulas;

$$\boldsymbol{Horizontal\ Gradient}:\ \boldsymbol{g_x(x,y) = I(x+1,y) - I(x-1,y)} \tag{3.4}$$

$$\boldsymbol{Vertical\ Gradient}:\ \boldsymbol{g_y(x,y) = I(x,y+1) - I(x,y-1)} \tag{3.5}$$

$$\boldsymbol{m(x,y) = \sqrt{g_x(x,y)^2 + g_y(x,y)^2}} \tag{3.6}$$

$$\boldsymbol{\theta(x,y) = \tan^{-1}\left(\frac{g_x(x,y)}{g_y(x,y)}\right)} \tag{3.7}$$

The entire image is split into cells, which are the spatial square regions with a pre-defined size of pixels. For each cell, HOG is computed by an angle and magnitude vote into histogram bins, which in turn are denoted as h_i where *i =1, 2, …,n* and *(hr,hθ)* denote the smallest possible interval, where angle *θ(x,y)* fits; lastly, *hr, hθ* histogram bins are defined as:

$$\boldsymbol{H(r) = H(r) + \frac{\theta(x,y) - h_r}{h_q - h_r} m(x,y)} \quad \boldsymbol{H(q) = H(q) + \frac{h_q - \theta(x,y)}{h_q - h_r} m(x,y)}$$

$$\boldsymbol{where\ h_r \leq \theta(x,y) < h_q} \tag{3.8}$$

After computing each cell histogram, all histograms are concatenated in a single vector. Because of the illumination variations in the image, cell histogram normalization should be done. For a group of cells called 'block', normalization is done according to the related factor defined as:

$$\boldsymbol{f = \frac{V}{\|V\| + \varepsilon}} \tag{3.9}$$

where $\boldsymbol{f}$ represents a normalized histogram vector, V is a non-normalized vector containing all histograms in a given block, $\boldsymbol{\|V\|}$ is the L2-norm of vector V and $\boldsymbol{\varepsilon}$ is a very small constant, standing as the non-zero value for all histograms in a given block. Finally, the normalized vectors are stacked together into the feature vector.

3.3 Classification

3.3.1 Support Vector Machine (SVM)

Pavnik and Cortes [105] first proposed a classification method which linearly classifies the hyper plane optimally. The idea behind the support vector machine (SVM) is to separate the largest fraction points by maximizing the distance of each class from the hyper plane. As shown in Figure 3.8, the maximal margin between the vectors of the two classes are separated with an optimal hyper plane.

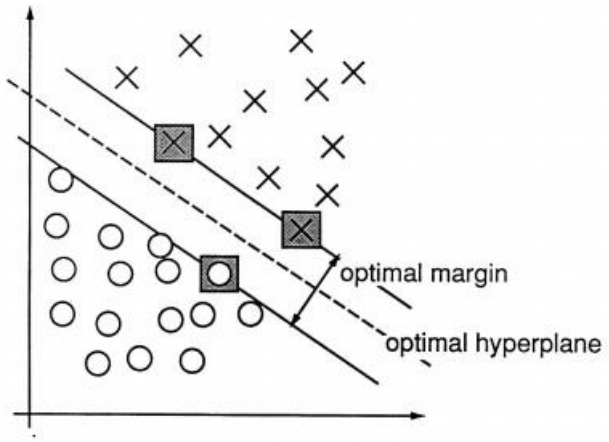


Figure 3. 8. An example of a separable problem in a 2-dimensional space. The support vectors, marked with grey squares, define the margin of largest separation between the two classes. [105]

Especially for two-class identification, SVM is a very effective classifier; hence, it can be successfully applied for gender classification problems [53]. The test image in the SVM classifier is identified based on the class to the class which has the maximum distance to the closest point in the training. The SVM training algorithm builds a model, which estimates if the test image is identified as male or female. To determine the gender in the image correctly, SVM needs a considerable amount of training data in order to select an effective decision boundary.

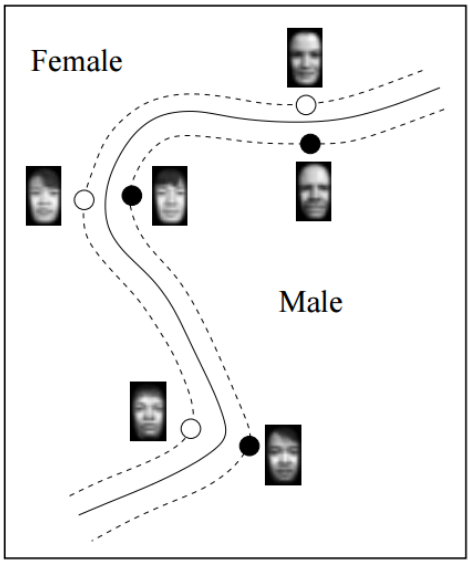


Figure 3. 9. Support faces at the boundary [53]

In Figure 3.9, three male and three female support faces from an actual SVM classifier is shown. The optimal separating hyper plane and its margin, shown as a dashed line, are depiction of gender classification with SVM.

SVM is a binary classifier which separate the hyperplane into two regions linearly. However, every problem cannot separate hyperplane linearly, and non-linear functions used to separate hyperplane into two regions. In order to classify nonlinear problems, kernel tricks are used such as: gaussian (radial based), polynomial, quadratic, multi-layer perceptron. The idea is optimizing the features according to the input space independent from the coordinates, in order to maximize the margin and make the features separable by mapping the instance in higher dimensional space.

Radial Based (Gaussian) Kernel: This kernel generally used for image-based problems and the function is defined as;

$$K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2), \tag{3.10}$$

where x_1 and x_2 are two samples of feature vectors.

Polynomial Kernel: Function definition is

$$K(x_1, x_2) = (1 + x_1' x_2)^p, \tag{3.11}$$

where $x_1 \; and \; x_2$ are two samples of feature vectors and p is a positive integer.

Multi-Layer Perceptron: Function definition is

$$K(x_1, x_2) = tanh(p_1 x_1' x_2 + p_2), \tag{3.12}$$

where $x_1 \; and \; x_2$ are two samples of feature vectors, p_1 is a positive integer and p_2 is a negative integer.

3.3.1.1 Multi-SVM Classification for Age Prediction

SVM is the classification method to distinguish two different classes. On the other hand, for multi-label classes there are two methods to classify multi classes; one-versus-all and one-versus-one [106] [107].

One-versus-all: The idea of the multi-label SVM for one-versus-all method is to compare a class with all other classes to see if the item belongs to any of the classes or not. If the class is not assigned to one, then other classes also tested individually. Let $D = \{(x_1, y_1), \dots (x_m, y_m)\}$, where $x_i \; (x_i \in R^n)$ denote an n-dimensional feature vector and $y_i \; (y_i \in Y_p)$ denote a *p*-dimensional label vector which has {-1,+1} implying that +1 shows the data belonging to this class, while -1 shows that the data does not belong to this class. The size of the label set is denoted with *p*. The *j*-th component of y_i denotes the output of *j*-th binary SVM. For the dedicated *D*, using the binary SVM classifier, each class is separated from the other classes. As a result, the *p* binary SVM classifiers are formed entirely [108].

One-versus-one: Another multi-label classification method is one-versus-one and the idea is to combine a couple of classes to binary classification among all classes. Then, the most accurate one is grouped into the corresponding class. Let $D = \{(x_1, y_1), \dots (x_m, y_m)\}$, where $x_i \; (x_i \in R^n)$ denotes a n-dimensional feature vector and $y_i \; (y_i \in Y_p)$ denotes a *p*-dimensional label vector, which has {-1,+1}, implying that +1 shows the data belongs to this data and that -1 shows the data belongs to another class while the rest of the classes are ignored. The size of the label set is denoted with *p*. The *j*-th component of y_i denotes the output of *j*-th binary SVM. For the design *D*, using the binary SVM classifier, each class is separated from each pair of other classes.

3.3.2 K-Nearest Neighbor Classifier (KNN)

K-Nearest Neighbor (KNN) classifier is the most widely used classification and prediction method and an example of instance-based learning, where the training data is stored first and the test data is classified through comparison with the most similar data in the training set [109]. The similarity of the data is measured by distance function, which is most commonly chosen as the Euclidean Distance:

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{3.13}$$

where x_i=x_1, x_2, x_3… x_m and y_i=y_1, y_2, y_3… y_m denotes the m attribute values of two records. When k=1, the test data is assigned to the shortest distance. If the k values are greater than 1, the KNN method becomes sensitive to outliers. In this way, the outliers' effect is reduced and the distribution is smoothened [110].

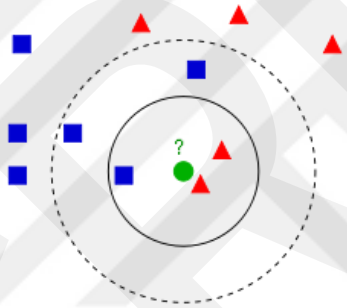


Figure 3. 10. Example of KNN classification [111].

In Figure 3.10, the green circle is the test point to be classified to the first class (blue squares) or to the second class (the red triangles). If k=3, the inner circle should be considered, there are two red triangles and one blue squares; thus, the green circle is identified as a red triangle. If the k=5 outer-dashed circle is considered, there are three blue squares and two red triangles, while the green circle is classed as a blue square.

For gender classification from facial images, KNN is an appropriate choice to use because of its efficiency. Since this classification needs less memory storage and compatibility, it has satisfactory discriminative properties, not to mention its efficiency for image distortions such as; rotation, illumination etc. problems [112].

3.4 Artificial Neural Network (ANN)

The origin of the neural networks is the working principle of the human (or animal) brain. The structure of neurons in the brain and their working principles are imitated for learning. As the human brain has dendrites, cell body, axon, and synapse parts, an

electronic implementation of such neural networks has the inputs which correspond to the dendrites, the neuron part which corresponds to the cell body and the axon and, lastly, the outputs correspond to the synapses. In the neuron parts, there are weights to multiply by each input. All the products are summed with biases and the sum is sent to the transfer function. The simple representation is shown in Figure 3.11.

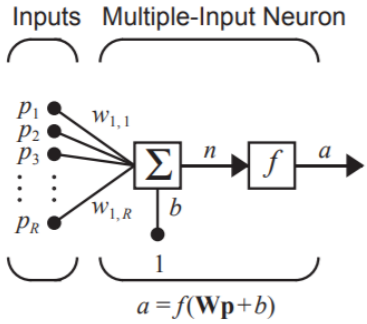


Figure 3. 11. Multiple Input Neuron [113]

As it is seen in Figure 3.11, there are R inputs shown as $p_1, p_2, \dots, p_R$, and their corresponding weights are $w_{1,1}, w_{1,2}, \dots, w_{1,R}$, while b is the bias and n is the sum of all weighted inputs and bias as shown in the following equation.

$$n = w_{1,1}p_1 + w_{1,1}p_1 + \cdots + w_{1,R}p_R + b \tag{3.14}$$

Because the inputs and weights can be written in matrix form as w_p, the output of the transfer function a, which is also the output of the neuron, can be written as the following equation form.

$$a = f(w_p + b) \tag{3.15}$$

This representation is used only for one neuron and one layer. For our case, the number of neurons and layers are more than one. Consequently, the corresponding neural network representation appears as in Figure 3.12.

In Figure 3.12, R is the number of inputs, S^1 is the number of neurons in the first layer, and S^2 is the number of neurons in the second layer. The number of neurons in each layer is different from the others. Also, the output of the first layer (a^1) is the input of the second layer, while the output of the second layer (a^2) is the input for the third layer; the last output (a^3) is the output of the neural network. For this reason, the third layer is called the *Output Layer* and the first and second layers are the *Hidden Layer* of the Neural Network. Such networks are called *Multilayer Neural Network* and are more effective than single-layer networks. Because single-layer networks have only one transfer function, they are not complex and not so detailed with respect to multi-layer neural networks. There are also more complex neural networks with 50, 100 or

more layers called *Deep Neural Networks.* For the purpose of the present research, multilayer neural network is used for body features and deep neural network are used for face images.

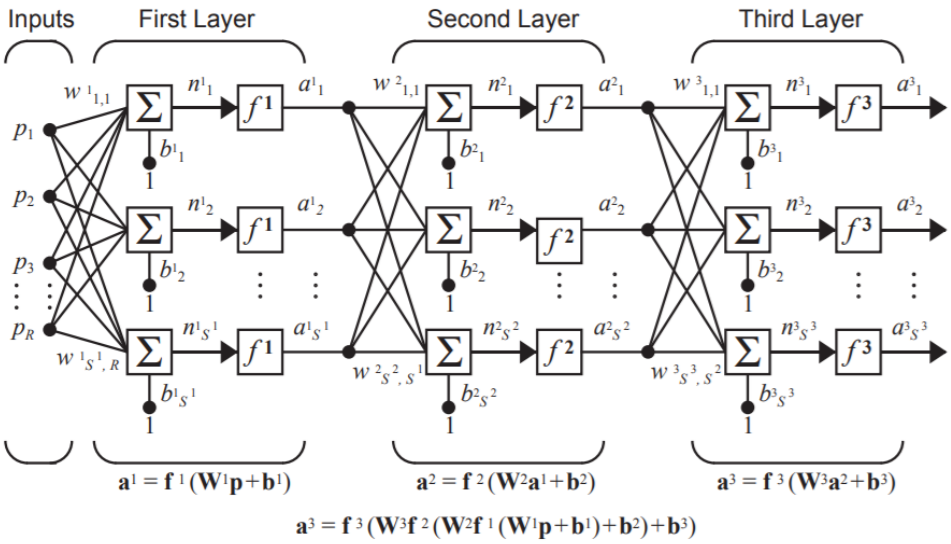


Figure 3. 12. A Three-Layer Network [113]

For the neural network, the transfer function is also an important part of the network and the sigmoid, sine, hyperbolic tangent, etc. functions are mostly used for this purpose.

Last but not least, the artificial part of the neural network is the training phase, which provides learning according to the target. In turn, the learning process occurs by updating the weights of the inputs and the biases of the network. This learning process is called the *Learning Rule* and is used to train the neural network in order to adjust the weights and biases to converge to the target output. To define the learning rule, the following formulas are used in general.

$$w^{new} = w^{old} + ep^T \tag{3.16}$$

$$b^{new} = b^{old} + e \tag{3.17}$$

$$where,\ e = t - a \tag{3.18}$$

In the above equations, W denotes weight (old/new) vector, b denotes the biases (old/new), p^T denotes the transpose of input vector, t denotes the target output, a denotes the actual output, and e denotes the error, which is the difference between the target and actual output. While converging to the target output, all the weights, biases and errors are handled together with different optimization algorithms, such as Levenberg-Marquardt, Gradient Descent [114], Bayesian Regularization [115], BFGS

Quasi-Newton [116], Scaled Conjugate Gradient [117], Resilient Backpropagation [118], etc.

3.5 Convolutional Neural Network (CNN)

In 1989, LeCun et al. [119] proposed the first Convolutional Neural Network (CNN) for the recognition of handwritten zip codes used in real world problems. They constructed the LeNet-5 network which, at the time, was limited due to resources for computation since training longer networks was a challenging task as regards the algorithms. Although CNN has more potential due to its high number of neuron layers, it has become even more popular recently thanks to the formation of Graphical Processing Unit, increase in images (datasets) available on the Internet to train networks and, finally, more effective algorithms developed to deal with complicated problems. A long time later, AlexNet [120] was initiated, reviving CNN with 8 layers and significantly improving it in such a way that the next system, ResNet[121] came out with 152 layers. Today, there are quite well-known networks, such as VGGNet [122], GoogleNet [123], etc., each with a different and much higher number of layers. There is a wide range of areas for use with CNN, especially in computer vision (face recognition [124], image classification [125], action recognition [126], human pose estimation [127]), natural language processing (speech recognition [122] and text classification [128]) [129]. The reason for this is the power of the subject-independent feature selection mechanism, which has not a pre-processing step and, instead, is included in the mechanism.

A CNN is a multi-layer neural network with special layers, namely convolutional layer, non-linearity layer, pooling layer, and fully connected layer. The structure of a CNN can be shown as below:

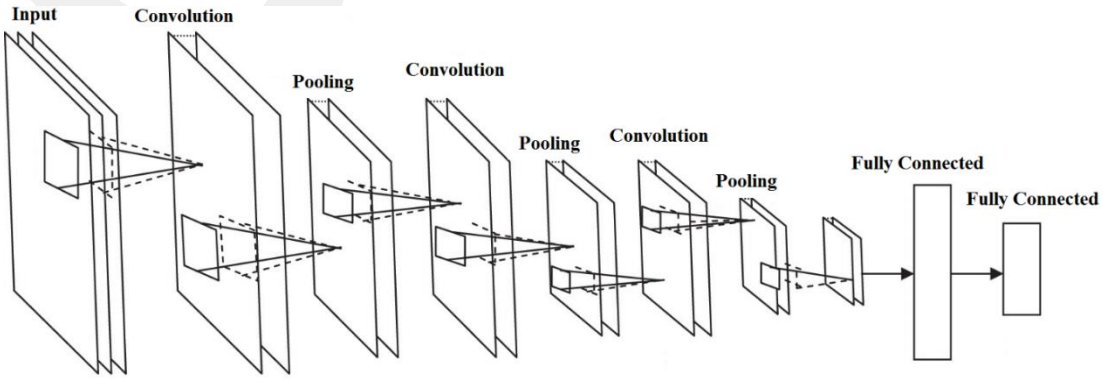


Figure 3. 13. The General Structure of a Convolutional Neural Network [130]

In what follows, we will look at each layer in detail.

Convolutional Layer: The main layer of CNN which processes the convolution operations with kernels is also called a 'convolution filter'. Convolution operations are generated with a set of kernels on the receptive fields of the input image to construct the feature map. All the feature maps generated from different filtering operations are combined to form the layer.

Non-Linearity Layer: Possessing non-linearity is important for multi-layer neural networks in order to be powerful. Such non-linearity is provided by passing the weighted sum of its inputs using the activation functions which are generally sigmoid, tanh and ReLU (Rectified Linear Units). ReLU [131] function is crucial for training time which is much shorter than other activation functions for running with the real capacity of network.

Pooling Layer: Generally, the pooling layers are used after the Convolutional Layer to reduce height and width, but not depth. The purpose of using a pooling layer is to increase the speed of training time and decrease or prevent overfitting. In this layer, a feature map down-sampling is generated by taking the maximum value of the pooling window so as to maintain the important parts of the feature map. Still though, other functions, such as average and L2-norm pooling, may also be used.

Fully Connected Layer: In general, there are a couple of fully connected layers to conclude the CNN after convolution and pooling operations. These layers are simple ANN structures and their input is the one-dimensional conversion of the previous pooling layer.

Softmax Layer: In neural networks (NNs), the most commonly used activation function is sigmoid. However, the sigmoid is used only for two-class output NNs, is not unique, and does not give any information regarding back-propagation in the training stage. For NNs with more than two outputs, Softmax is used as the activation function to determine the probability of each class. The sum of all outputs' probabilities equals to 1.0, yielding the Softmax function equation as:

$$\sigma(z)_i = \frac{e^{z^j}}{\sum_{k=1}^{K} e^{z_k}} \tag{3.19}$$

where z is the vectors as the inputs of the output layer, K is the output units and j ($j=1, 2, \ldots, K$) is the indexes of K.

3.6 Cross Validation

Cross validation is an evaluation method of a model to generalize the result for an independent data set. To estimate the performance or accuracy of the classification or prediction, data set is divided into training and testing parts and the evaluation is repeated according to the data division technique. This process provides more results for different combinations of data set in limited number of training and testing sets. The mostly known cross validation techniques are *k-fold* and *Leave One Out*.

3.6.1 k-fold Method

In this technique, data set is divided in to *k* sets with equal size. *k-1* sets are used for training the system and the rest one is used to validate the system. This process is repeated *k* times and each time with different validation set. So, each set is validated exactly once. Then the accuracy result is calculated by averaging of all *k*-validation. Generally, 10-fold cross validation is used as in this thesis.

3.6.2 Leave One Out Method

In this technique, one observation in a data set is separated as validation set and the rest of the observation in the data set are separated as test set. Then this process is repeated for the number of observations in the data set. Each observation is validated exactly once and the result of estimation is calculated by averaging of all validation. Leave one out method is more expensive but more reliable than k-fold method.

3.6.3 Leave One Person Out Method

This technique is used for this thesis to validate each person and similar to leave one out method. In this thesis, there is limited number of people (which is 170) in order to find the accuracy of gender and age prediction for body information. Therefore, data of one person is used as validation set and the rest of the data is used as test data. This process is repeated 170 times which is the number of candidates in this study.

CHAPTER 4

RESULTS OF GENDER PREDICTION

A main purpose of the present study, as stated earlier, is automatic gender prediction. There are two approaches to predict gender: face based and body-based. Firstly, the experimental results of the face-based approach are presented. Next, the body-based approach is clarified with the results. In the face-based approach, statistical pattern recognition methods are applied, followed by the body-based gender prediction approach involving two methods: the statistical pattern recognition approach given in Section 4.2, and body-based artificial neural network in Section 4.3.

4.1 Gender Prediction from Face Images Using Statistical Methods

The initial state of this study is gender prediction from face images. Therefore, the statistical methods in Section 3.2 and 3.3 are used to predict gender from face images. This section shows the experimental results of these statistical methods.

The Facial Recognition Technology (FERET) database [55] is mostly used in gender prediction studies, offering 1262 images in total with 531 female and 731 male images from the frontal face. The original sizes of the images are 256×384. The FERET database is first used both for the training and testing phase of the classification; then, the Vital Longevity Face Database, created at the University of Michigan, is utilized to test the images with LBP and HOG feature extraction methods. In the database, there are 575 individual faces ranging from ages 18 to 63 [96]. In the part related to gender classification, a total of 153 gray-level images are used with 93 females and 60 males, with the original sizes at 646×480, but resized to 128×192.

4.1.1 LBP Gender classification Result

For gender classification – and as stated before - LBP is the mostly used feature extraction method. In our experiments, SVM and KNN classification methods have been applied to the FERET database images. SVM is applied in the Linear kernel function, whereas the KNN classifier is used with k=1, 2, …, 20 values. As shown in Figure 4.1, the best KNN result is 77.08 % accuracy with k=6. However, SVM with Linear kernel function achieves 83.19 % accuracy, which is better than the best accuracy of KNN classifier within the present work. For the value of k, from 1 to 20, the result of gender predictions is shown Figure 4.1.

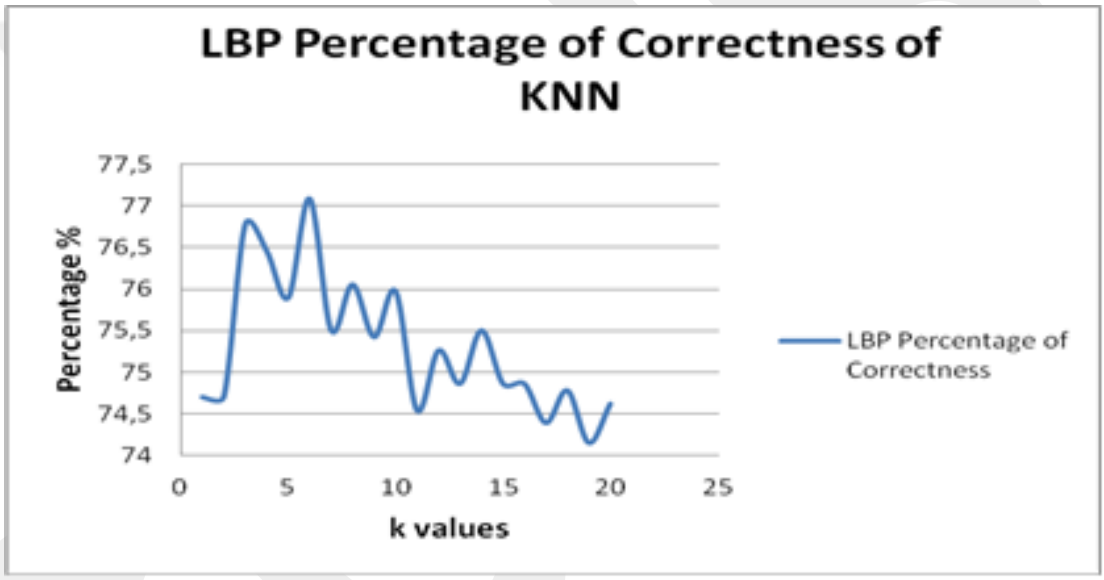


Figure 4. 1. LBP accuracies for the k values of KNN classifier

To predict the gender of those images at the FERET database, the elapsed time is calculated throughout 30 images tested for SVM and KNN classifiers, resulting in 9.01 seconds for SVM and 6.90 seconds for k equal to 6 in KNN. The time requirements to process 30 images is considered because in one second, 30 images are captured by a camera in actual settings. To calculate the elapse time, only the best accuracy conditions are examined.

4.1.2 HOG Gender classification Result

Another feature extraction method is HOG for gender classification. SVM and KNN classification methods are applied after using the HOG feature extraction method on the FERET database. The same processes applied to the LBP method are applied to

the HOG method. In addition, SVM is used with the Linear kernel function, while KNN is used with different k values from 1 to 20. This time, after k=12, the accuracy of the gender prediction is nearly stabilized at 85 %, as seen in Figure 4.2.

As can be seen, the SVM test result with 88.74% accuracy is more powerful than the best value of KNN classifier with 86.68 % by k=8.

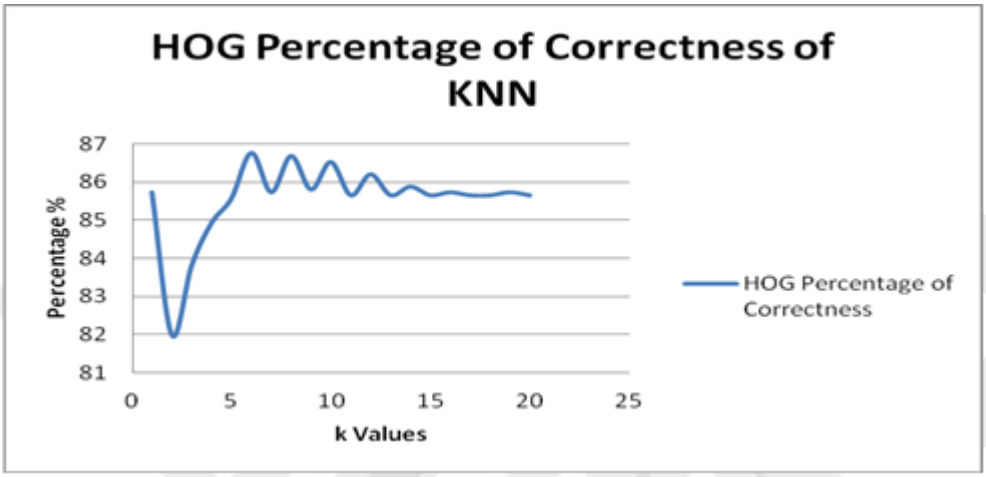


Figure 4. 2. HOG accuracies for k values of KNN classifier

For HOG feature extraction, the best accuracy values are obtained in 0.26 seconds using the SVM classifier, and 0.22 seconds using the KNN classifier, to classify the gender of people in 30 images.

4.1.3 Comparison of Feature Extraction and Classification Methods

For the same method of classification, SVM is more powerful than KNN in two feature extraction methods: LBP and HOG. According to the comparisons, the accuracy and time of HOG is better than LBP as shown in Table 4.1. HOG accuracy levels are greater than LBP, while the required time to process 30 images by HOG is also less than LBP.

Table 4. 1. Best accuracy levels in gender prediction for LBP and HOG feature extraction method and SVM and KNN classifiers, and time requirements to process 30 images

	LBP		**HOG**	
	Accuracy	**Time(s)**	**Accuracy**	**Time(s)**
SVM	83.19 %	6.90 s	88.74 %	0.26 s
KNN	77.08 %	9.01 s	86.68 %	0.22 s

For the KNN classifier, the accuracy value of LBP is lower than that of HOG, as shown in Figure 4.3. Additionally, stabilization for HOG accuracy can be observed after k=12 although LBP's accuracy value is not stabilized in any k value.

Up to this point, all experiments have been done on the FERET database using the *Leave One Out* technique. Additionally, the gray-level images of the Vital Longevity Face Database were also used for test images, and the FERET dataset was used to train images, all resized to 128 X 192 in order to have unity in the feature matrix.

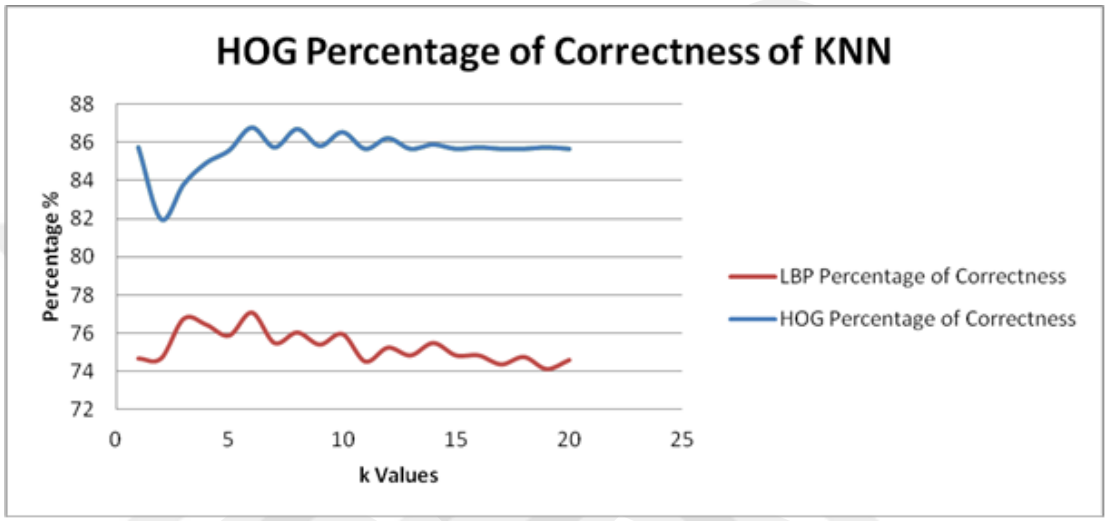


Figure 4. 3. LBP and HOG accuracies for k values of KNN classifier

Apart from these feature extraction methods, Lin et al. [101] proposed that the combination of LBP and HOG gives better result in gender prediction, having used *t-test* to reduce feature dimensions. The accuracy result for LBP is 86.75%, HOG is 87.52% and LBP+HOG is 91.50% according to their results. For this reason, we also use a combination of LBP and HOG in our experiments. The accuracies of the gender classification for LBP, HOG and LBP+HOG feature extraction with SVM and KNN classification methods are shown below.

Table 4. 2. The accuracy of gender prediction for LBP, HOG and LBP+HOG feature extraction method and SVM and KNN classifiers

	LBP	**HOG**	**LBP+HOG**
SVM	50.95 %	63.69 %	67.97 %
KNN	42.67 %	59.87 %	56.86 %

As seen in Table 4.2, the accuracy of the SVM classifier is less than KNN in each different feature extraction method, implying that SVM is also more effective than KNN when used in different face databases other than FERET. Plus, the combination of LBP and HOG feature extraction methods had better performance than their use independently.

4.1.4 Working on Face Region of the Image

So far, images have been used after resizing and without any cropping. In this part, the face region of the images is cropped according to the Viola-Jones Algorithm. Later, they are resized to 161 X 121 in order to obtain unique number of features; in addition, LBP, HOG and LBP+HOG feature extraction methods are applied on the Vital Longevity Face Database. Lastly, SVM is used to classify gender prediction by the *Leave One Out* method.

Table 4. 3. Accuracy of the gender prediction with LBP, HOG and LBP+HOG feature extraction method and SVM classification method are applied only to face region

	LBP	**HOG**	**LBP+HOG**
Accuracy	81.69 %	86.92 %	89.54 %

Table 4.3 shows that the combination of two feature extraction methods (LBP and HOG) also provides the best result for the face region. Furthermore, and considering only the face region for feature extraction, the classification accuracy is increased from 84.31% to 89.54 % using the same size (161 X 121) and extraction method (LBP+HOG).

4.1.5 Gender Prediction Using Statistical Methods- Conclusion

So far, the study of gender prediction has been explained step by step and in detail. The proposed methods, LBP and HOG for feature extraction and SVM and KNN classification are defined briefly. Then, the results of the experiments with the proposed methods were shown with tables and figures. Results show that SVM is a better classification method than KNN. Two face databases, FERET and the Vital Longevity Face, are used with the *Leave One Out* technique first and, then, one of

them was used for training and the other one as test images. It is concluded that only a database with the same type of images should be used for high accuracy.

Finally, because related studies [101], [103], having used LBP, HOG and SVM methods, worked only on the face region in the images, the same approach was adopted in the present work. In Lin's study [23], t- a t-test is used to decrease the number of features, yielding the best accuracy with LBP+HOG+t-test at 92,21% on the FERET database images. In Singh's study [25], the LBP and HOG methods are applied separately and the best accuracy is 95.56% on the Indian Face Database (IFD). As a result, the LBP+HOG feature extraction method with SVM classification is shown to offer the best accuracy (89.54%) on the face region in the Vital Longevity Face Database images.

4.2 Gender Prediction from 3D Anthropometric Measurements Using a 3D Camera

In this part of the study, an automatic gender detection system is proposed using the anthropometric measurements of the human skeleton. These values refer to the size- and shape-related descriptors of the human body and are used to provide measurements in several applications such as garments, equipment and application materials, ergonomics and architecture. The measurements include the height, width and weight of the human body and its parts [21]. Since body proportions are different for men and women, the coordinates of joints provided using motion-tracking technology allow one to find the individual length measurements and proportions of their human subjects [22]. Our approach is based on determining the lengths between the joint points and height of the people. To verify this approach, the accuracy for gender detection is experimentally tested using 60 subjects in different age ranges. The 3D coordinates of joint points are obtained using a 3D camera. To predict the gender according to the features obtained by the lengths between the joints and the height of each person, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and Support Vector Machine (SVM) are used.

4.2.1 Data Collection

As no alternative datasets could be found containing both the skeleton joint position data, or any other datasets large enough with gender information, we collected our own dataset. Sixty (60) volunteers (29 - female, and 31 – male, all aged 20 to 60) participated in the experiment. The test environment, shown in Figure 4.4, is located in one of the laboratories at Atilim University, Ankara, Turkey. The KinectV1 is used to record the body metrics from volunteers.

Each volunteer was asked to stand at ease in front of the camera, free to take any position or do any action for any period of time. There was no restriction on time because the average values of all frames were taken for each target to extract the features and generate the necessary vectors.

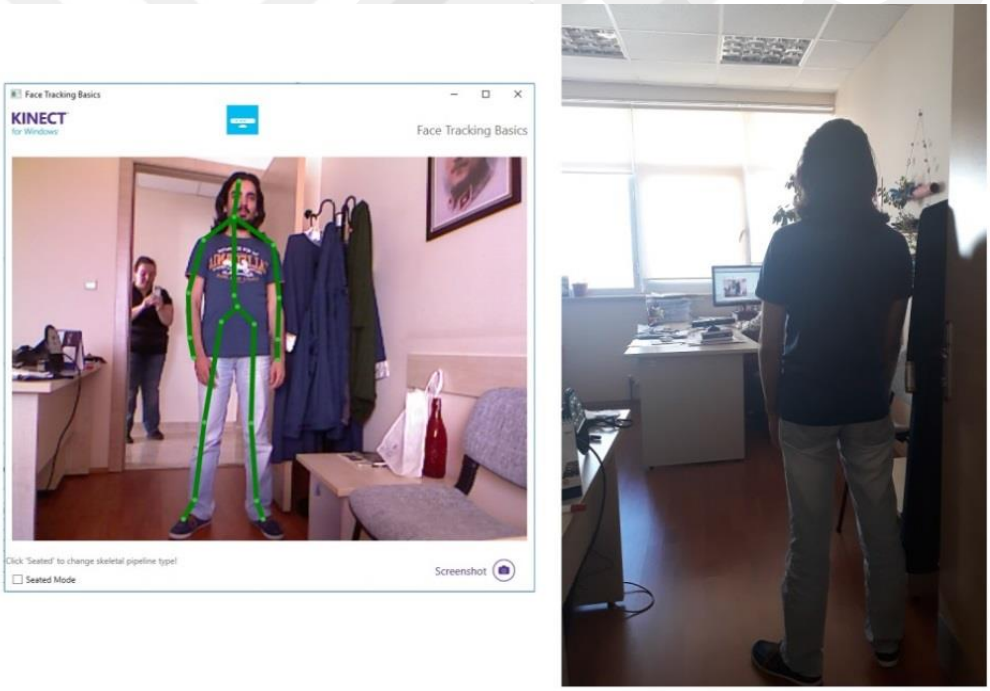


Figure 4. 4. Test Environment at Atilim University.

4.2.2 Origin of the Idea

To predict the gender, differences between men and women have so far been investigated in many studies, some of which have been discussed earlier in this thesis. For our problem, body proportions and height of each individual is considered to predict their gender. The research proposed by Loomis [132] gives the ideal proportions in centimeters for both genders as shown in Figure 4.5.

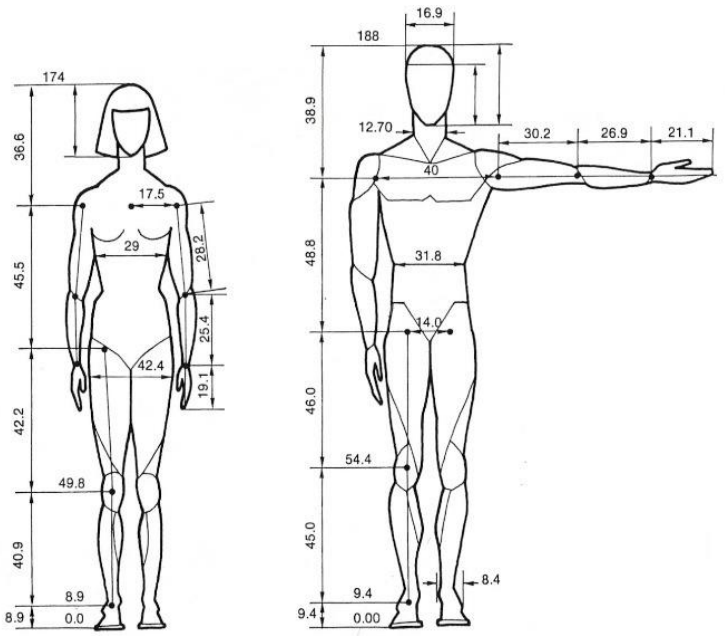


Figure 4. 5. Ideal gender proportions [132] in centimeters: female (left) and male (right).

In Figure 4.5, the left illustration corresponds to females and the right one to males, and the proportion of each part of the body is numerically represented in centimeters. We see that for men and women the proportions of some parts of the body are different. For instance, women's hips are larger than men's, while men's shoulders are larger than women's. In [72], these lengths are used with the height of humans to classify the images according to gender. In this study, similar to the approach mentioned above, we assume that the distances between the joint points might change depending on the gender of a person. Therefore, we obtain the positions of the following joint points and perform feature extraction from them.

4.2.3 Obtained Joint Positions and Features

In this part of the study, we use a Microsoft Kinect 3D camera, which is a motion-capturing device that not only acquires the standard RGB image of the object, but also the distance (or depth information) between the object and the camera. The camera has five components: infrared (IR) Emitter, Color Sensor, IR Depth Sensor, Tilt Motor and Microphone Array.

To acquire the depth image and information, the IR Emitter emits IR light beams while the IR Depth Sensor scans the reflected beams back, and the depth information is measured as the distance between the sensor and target.

The Kinect skeletal model provides 20 joint points of the person being tracked with the depth information [133] as follows: hip center–1, spine–2, shoulder center–3, head–4, shoulder left–5, elbow left–6, wrist left–7, hand left–8, shoulder right–9, elbow right–10, wrist right–11, hand right–12, hip left–13, knee left–14, ankle left–15, foot left–16, hip right–17, knee right–18, ankle right–19, and foot right–20 (see Figure 4.6).

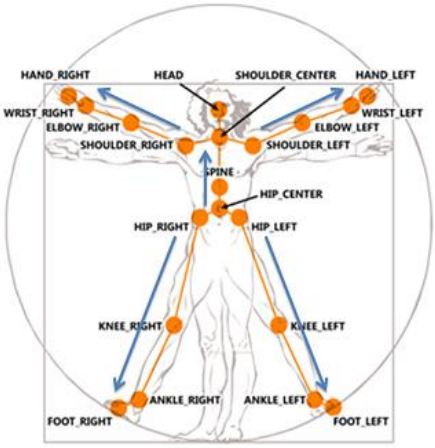


Figure 4. 6. Skeleton Position Joints obtained by Kinect v1 (adopted from [134]).

These 20 joint points are used to obtain the features used in the classification stage. The features include the Euclidian distance between any two joint points and the height of the target. If needing a formula, the position of the joint is shown as p_i, defined as follows:

$$p_i = (x_i, y_i, z_i) \quad where\ i\ \in \{1,2,3,\dots,20\} \tag{4.1}$$

The Euclidian distance between the joints is defined as:

$$distP_{m,n} = \sqrt[2]{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2} \tag{4.2}$$

where m and n are two different joint points numbered from 1 to 2020, $m, n \in \{1, 2, \dots, 20\}$ Mentioned joint points are listed above with their numbers.

The distance between each joint $distP_{m,n}$ with any other joint is calculated using Eq. 4.2. There are 190 binary combinations of twenty joint points; hence, 190 features. Additionally, three more features (height by skeleton, height by estimation and height by wingspan) are calculated as given by Eq. 4.3 & 4.4 [72].

Height By Skeleton: The total length from foot to head plus half of the head height, because the head position coordinates are driven from the center of the head.

$$\text{Height by Skeleton} = L_{H\text{-}SC} + L_{SC\text{-}S} + L_{S\text{-}HC} + L_{HC\text{-}K} + L_{K\text{-}A} + L_{A\text{-}F} \quad (4.3)$$

where $L_{H\text{-}SC} = distP_{\text{head}-\text{shoulderCenter}}$, $L_{SC\text{-}S} = distP_{\text{shoulderCenter}-\text{spine}}$,

$L_{S\text{-}HC} = distP_{\text{spine}-\text{hipCenter}}$, $L_{HC\text{-}K} = distP_{\text{hipCenter}-\text{kneeLeft/kneeRight}}$,

$L_{K\text{-}A} = distP_{\text{kneeLeft/kneeRight -ankleLeft/ankleRight}}$,

$L_{A\text{-}F} = distP_{\text{ankleLeft/ankleRight}-\text{footLeft/footRight}}$.

Height By Estimation: This height calculation method depends on the idea that the whole body is not captured by camera, leaving out some of the upper or lower parts of the body. The proposed method in [132], which recommends that the distance between the shoulders and knees is 52% of the overall height according to the ideal body proportion, is applied for height estimation.

Height By Wingspan: This height calculation method depends on the idea that the subject is considered in a sitting position. According to the ideal body proportions (Figure 4.5), the length of the wingspan equals the height of the target.

$$\text{Height by Wingspan} = L_{HL\text{-}WL} + L_{WL\text{-}EL} + L_{EL\text{-}SL} + L_{SL\text{-}SC} + L_{SC\text{-}SR} + L_{SR\text{-}ER} + L_{ER\text{-}WR} + L_{WR\text{-}HR} \quad (4.4)$$

where $L_{HL\text{-}WL} = \text{distP}_{\text{handLeft-wristLeft}}$, $L_{WL\text{-}EL} = \text{distP}_{\text{wristLeft-elbowLeft}}$,

$L_{EL\text{-}SL} = \text{distP}_{\text{elbowLeft-shoulderLeft}}$, $L_{SL\text{-}SC} = \text{distP}_{\text{shoulderLeft-shoulderCenter}}$,

$L_{SC\text{-}SR} = distP_{\text{shoulderCenter}-\text{shoulderRight}}$, $L_{SR\text{-}ER} = distP_{\text{shoulderRight}-\text{elbowRight}}$,

$L_{ER\text{-}WR} = distP_{\text{elbowRight}-\text{wristRight}}$, $L_{WR\text{-}HR} = distP_{\text{wristRight}-\text{handRight}}$

Together with these three features, a total of 193 features are obtained for each frame. In order to avoid any unbalanced feature number depending on the frames and missing joints may obtain by emitting the reflected IR light beams from the Kinect camera, the average values of each feature are calculated for all frames to prepare the feature vector for the classification stage.

While the data was being collected, the actual heights of the individuals were also recorded to verify our height calculations, we took a meter ribbon and measured each one manually and the ability to capture the data of Kinect camera. When the actual and

experimental results are compared, it seems ±2.5 centimeters difference between each other.

4.2.4 Feature Selection Methods

In the classification stage, all features are used both individually and jointly for each classification method., though with possibly lower accuracy rates when used for classification. As testing each feature with different combination sizes for 193 features would be a very time-consuming process, not all combinations of 193 features were tested. On the other hand, the features should be combined based on a specific criterion to increase the classification accuracy. Such a criterion is determined according to the accuracy rate of each individual feature performance. Starting from feature 1 to 193, each accuracy rate of classification is checked. If greater than 70%, the individually classified feature is selected to combine other selected features; otherwise, the feature is not added to the selected features set. After the selected feature set is generated, an exhaustive search [135] is carried out to combine the selected features that satisfy the criteria, thereby testing all possible combinations for the KNN and SVM classifiers. For the neural network, feature selection is not applied because of its structure of weighted neurons.

4.2.5 Classification Methods

In this study, we used KNN (previously used for gender recognition in studies [136], [137]), SVM and Artificial Neural Network (ANN) to predict gender. These approaches are summarized in Section 3.3. In addition, linear, quadratic, polynomial, radial basis function (RBF) and multilayer perceptron function (MLP) kernels are applied here.

For the description and formal definition of the methods, we referred to [138]. As stated before SVM has also been used for gender recognition in previous studies [139], [140]. Also, ANN is widely used in pattern recognition and gender detection [141], [142] because of its satisfactory performance in applications. The ANN structure is formed by layers composed of individual neurons, each of which is associated with its weight learned during the training stage to reduce network error. In ANN, there are three layers, namely: input layer, hidden layer, and output layer. The input layer is the first layer where each sample of the network is used as an input. The hidden layer is

the second layer where error reduction is performed in the network. The last layer is the output layer where the desired outputs are determined with the same number of neurons to represent each output separately.

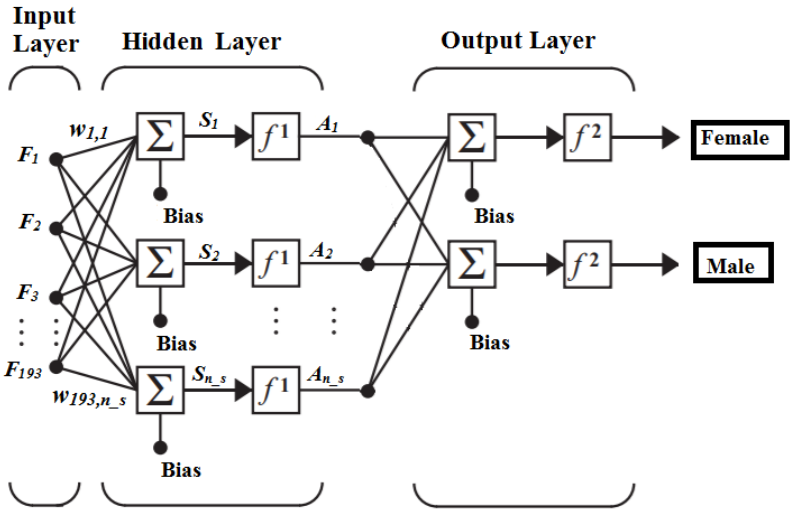


Figure 4. 7. ANN Structure of the Proposed System

Figure 4.7 shows the Neural Network representation used in this study. There are three layers: Input Layer, Hidden Layer, and Output Layer. In the input layer, F_1, F_2, ..., F_{193} denotes the all features and $w_{1,1}$, w_{193,n_s} denotes the weights of the related features, where w_{193,n_s} is the weight of 193 input features in n_s (where *n_s=10, 15, 20, 25, 30, 35, 40, 45* as the neuron size of Hidden layer. In the Hidden Layer, f^1 is the activation function, S_1, S_2, …, S_{n_s} are the weighted sums added to bias. A_1, A_2, …, A_{n_s} are the activation function results.

$$S_1 = w_{1,1}F_1+w_{1,2}F_2+\ldots+w_{1,193}F_{193}+Bias; \quad A_1 = f(S_1)$$
$$S_2 = w_{2,1}F_1+w_{2,2}F_2+\ldots+w_{2,193}F_{193}+Bias; \quad A_2 = f(S_2)$$
$$\ldots$$
$$S_{n\_s} = w_{n\_s,1}F_1+w_{n\_S,2}F_2+\ldots+w_{n\_s,193}F_{193}+Bias; \quad A_{n\_s} = f(S_{n\_s}) \qquad (6.5)$$

All steps are represented with the following algorithm.

The Study algorithm:

- For each volunteer, repeat the following process:
  - Capture all joint point coordinates from the 3D camera: $p_i = (x_i, y_i, z_i)$ *where* $i \in \{1,2,3,\ldots,20\}$

- For a joint point $m=1, 2, \ldots, 20$, calculate the distance between another joint point $n=1, 2, \ldots, 20$, *where* $n \neq m$: Using the equation (6.2)
- Calculate ***Height by Skeleton (HbS)***, ***Height by Estimation (HbE)*** and ***Height by Wingspan (HbW)***
- All calculated distances and heights are concatenated to create a feature vector for the person x where $x=1, 2, \ldots, 60$, $P_{xFeatures}$: $P_{xFeatures} = [distP_{1,1}, distP_{1,2}, \ldots, distP_{20,19}, HbS, HbE, HbW]$

➢ Each feature is considered individually and for each classification method found, the accuracy $Acc\_P_{ALL\_Features[x]}$, where $x=1, 2, \ldots, 193$

➢ If the accuracy $Acc\_P_{ALL\_Features[x]} \geq 70\%$, define the feature. x is the selected feature among 193 features.

➢ Among the selected features, different number of combinations of features are classified for each classification.

➢ The results are recorded, with the most accurate feature combination reported for each classification method.

4.2.6 Results and Discussion

To test the proposed approach, data was collected from volunteers and the obtained information from the 3D camera was processed to extract features used to classify gender. For this purpose, we used SVM with different kernel functions, KNN with different K values and ANN with three training functions (Scaled Conjugate Gradient, Bayesian regularization, and Levenberg-Marquardt optimization) applying the *Leave One Out* method to predict the gender of subjects.

Figure 4.8 shows the accuracy of classification using the KNN classifier. The black dashed line shows the accuracy when all the features are used, and the red dash-dot line shows the accuracy when the feature set is obtained by the exhaustive search.

The best accuracy achieved is 85.48% with K = 1 and K = 2 values while using the following feature combinations: Neck-Right Knee, Neck-Left Foot, Left Shoulder-Right Hip, Right Shoulder- Right Ankle. It should be noted that, as the k values increase, the accuracy tends to decrease as seen in Figure 4.8.

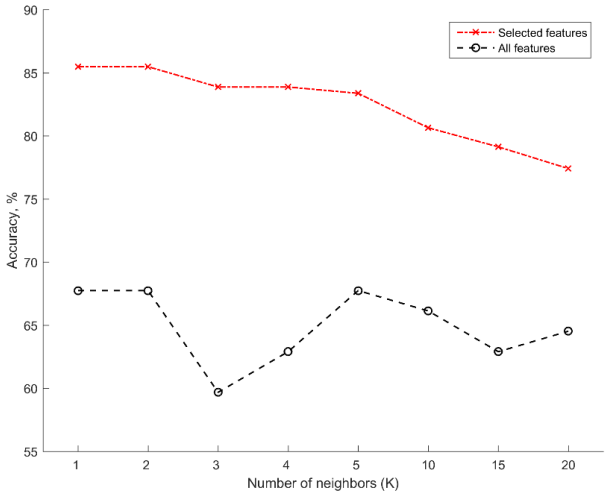


Figure 4. 8. KNN classifier accuracy results for gender detection vs different K values.

Table 4.6 presents the classification results of SVM classifier. The best accuracy is 96.77% while using the MLP kernel function with seven features (Head – Right Foot, Left Shoulder – Right Hip, Right Shoulder – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Knee – Right Knee, and Right Knee – Left Ankle). These features are illustrated in Figure 4.9. Distance between the cross joint points of feature combinations, which depend upon the width of the shoulder and the hip (identifiers for gender), gives the best accuracy for gender detection.

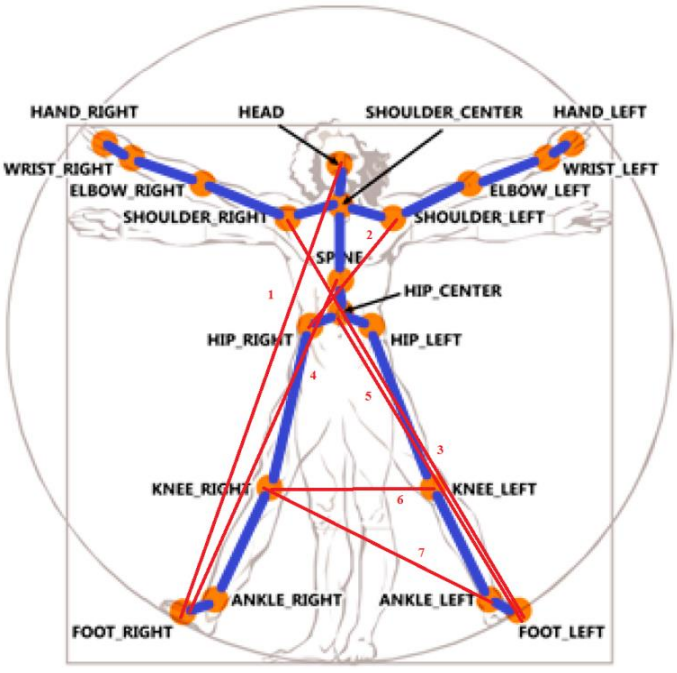


Figure 4. 9. Combination of features achieving best accuracy with MLP kernel function of SVM

As mentioned before, the best accuracy rates of the classification methods take different feature combinations as inputs. Not only the number of selected features, but also the sets of the selected features may be differentiated based on the classification methods In Table 4.6, the SVM classifier with Linear Kernel function achieves an 83.87% accuracy rate with twelve features used as inputs. Moreover, the SVM classifier with MLP Kernel function offers 96.77% as the highest accuracy rate. In addition, seven features, which do not exactly match the features used in the Linear Kernel function of SVM, have been used as inputs of the classifier. Other methods have been taken also a different number of selected features for the best accuracy. However, all possible combinations of the selected feature sets are obtained by testing every option. This process was time consuming. Hence, using other methods, such as Fisher Discriminant Ratio, would be better to select features that give the best accuracy. In future work, it is aimed to use Fisher Discriminant Ratio or similar feature selection methods.

Table 4. 4. Gender detection accuracies in SVM classification with feature combinations giving best performances.

Kernel Function	All Features	Best Accuracy	Feature Combinations Giving the Best Performance
Linear	63.33%	83.87%	Left Shoulder - Right Hip, Right Shoulder - Left Hip, Right Shoulder-Left Foot, Spine - Left Hip, Spine - Left Foot, Spine- Right Foot, Waist-Left Foot, Left Hip- Left Foot, Right Hip - Left Foot, Left Knee-Right Knee, Right Knee- Left Ankle, Right Knee-Right Foot
Quadratic	65.00%	93.55%	Head- Right Knee, Neck- Right Shoulder, Neck-Left Knee, Left Shoulder- Left Foot, Left Knee- Right Knee, Right Knee- Left Ankle
Polynomial	76.66%	93.55%	Neck-Right Shoulder, Left Shoulder-Right Hip, Spine-Left Hip, Left Knee-Right Knee
RBF	5.00%	93.55%	Head- Right Knee, Neck - Right Shoulder, Neck - Left Knee, Left Shoulder - Left Foot, Left Knee- Right Knee, Right Knee- Left Ankle
MLP	63.33%	**96.77%**	Head- Right Foot, Left Shoulder - Right Hip, Right Shoulder - Left Foot, Spine - Right Foot, Waist - Left Foot, Left Knee- Right Knee, Right Knee- Left Ankle

In Figure 4.10, the accuracy results of ANN are presented according to the number of neurons in the hidden layer, their training functions, and momentum. The training functions are defined to update the weight and bias values according to the Scaled Conjugate Gradient method, Bayesian regularization and Levenberg-Marquardt

optimization. The best accuracy result (76.66%) is obtained with the Levenberg-Marquardt optimization training function, 35 neurons in the hidden layer and 0,9 momentum value.

From the results of the experiment, we see that using the distance between each joint point with another one offers better accuracy than using the length of the two neighbouring joint points. The feature combination, which gives the best accuracy results, uses the distances between the joint points, which are not directly connected to each other.

The results obtained here are compared with those of Andersson et al. [15] and Sandygulova et al. [72], with the following differences detected:

Study [72] was performed in a 5-18 age group and the accuracy rate was 73%. In our study, the age range was between 20 and 60 years old and the accuracy is 96.77%. Study [15] achieved 95% accuracy using the gait and anthropometric measurements, while we have managed to achieve 96.77% accuracy using only the anthropometric measurements. In the latter case, furthermore, even when the camera cannot see the entire body, we can predict gender with the help of the ratio of the different joint points. Thereby eliminating the need for gait assessment.

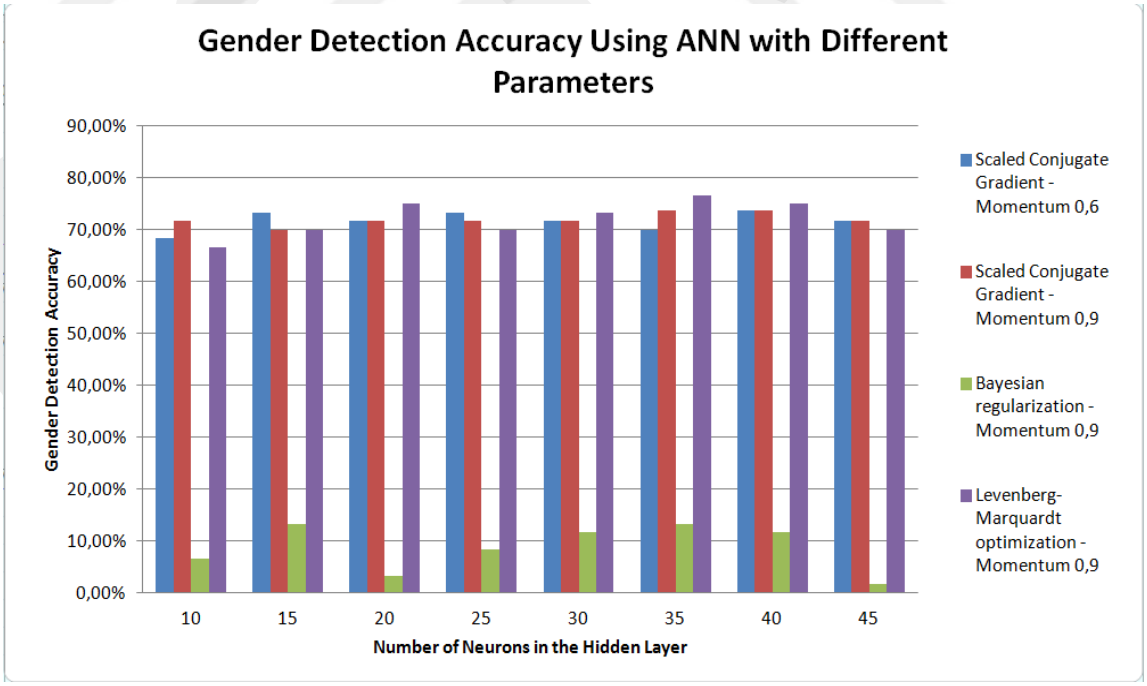


Figure 4. 10. Gender Detection accuracy using ANN with different parameters

While our results are still lower in accuracy than those achieved using face images (e.g., 95%, 97% and 98% in [48, 14, 49], respectively), our advantage is that we have

not needed high resolution face images of the person, and the gender can still be predicted if the face images cannot be acquired.

4.2.7 Gender Prediction from 3D Anthropometric Measurements using a 3D Camera - Conclusion

In this section, a new approach for gender detection has been proposed using both anthropometric body metrics and posture of people. Our main contribution, first, is that the anthropometric measurements of the person are automatically obtained by the developed system with the help of 3D camera. The second contribution is the unique dataset of skeleton joint positional data with subject gender information collected and made available.

For gender detection, we employed three classifiers (KNN, SVM, and Neural Network) with different parameters. On a dataset of records from 60 subjects aged from 20 to 60, the best performance for gender detection is 96.77% achieved with the MLP kernel of SVM classifier using eight selected features (Head – Right Foot, Left Shoulder – Right Hip, Right Shoulder – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Knee – Right Knee, and Right Knee – Left Ankle).

Also, the best accuracy is obtained with the combination of features, which does not directly depend on the width of the shoulders and hips, but is indirectly related with the widths. This approach is different from previous gender detection approaches by both considering the anthropometric and postural information rather than using the gait or the length of bones analysis approaches.

4.3 Gender Prediction from 3D Body Data Using ANN

One of the gender estimation methods we applied in section 4.2 was ANN, which had a lower performance than SVM and KNN methods. Considering that we have used ANN after the feature extraction step, it was assumed a more appropriate method to use the coordinates of the joint points obtained before feature extraction. Accordingly, we predicted the gender using the ANN classification method by first feeding the coordinates of all the joint points to the system and, then, using the coordinates of the lower and upper body joint points.

4.3.1 Data Obtained from the 3D Camera

As explained in Section 4.2.3, the image of the scene and coordinates of 20 joint points of each participant are captured using a 3D camera. These joint points are shown in Figure 4.6 and listed in Table 4.7. These 20 joint points are captured to track the body with limited computational resources and fluent interaction in real time [133].

Table 4. 5. Upper and Lower Joint Points of the Skeleton

Upper Body Joints		**Lower Body Joints**	
1.Head	7.Right Shoulder	11.Spine	16.Right Knee
2.Shoulder Center	8.Right Elbow	12.Hip Center	17.Left Ankle
3.Left Hand	9. Right Wrist	13.Left Knee	18.Right Ankle
4.Left Wrist	10.Right Hand	14.Right Hip	19.Left Foot
5.Left Elbow	11.Spine	15.LeftKnee	20.Right Foot
6. Left Shoulder			

In this part of the study, the points are captured and their 3D coordinates are stored as x, y, and z, each with 3 coordinates as 60 inputs in all for the neural network. Additionally, both the Upper and Lower Body Parts' joint points are considered as an input for the neural network. According to our consensus, the Upper Body Parts are described as the first 11 joint points, and the Lower Body Parts are described as the last 10 joint points, all in Table 4.7. Therefore, the input number for the Upper Body Part neural network is 33 (3x11) and for the Lower Body Part neural network, 30 (3x10). Only the joint point 'Spine' is used in both for the Upper Body Parts and the Lower Body Parts of the neural network because it is the center of the body.

The camera captures not only the joint points but also the images of the participants; and they are saved. The examples of different scene images are shown in the following figure which shows the participants' images in different environments.

In Figure 4.11, there are twelve images, six of which are a) Different Posture Examples, and the other six are the b) Different Places where the images were captured. In Figure 4.11-a, there are five different places where the images are captured. The first and the sixth images are taken in two different offices; whereas the second one is taken in a wide area, the third one is taken inside a meeting room, the fourth and the fifth are taken in two different houses.

Figure 4. 11. a) Different Posture b) Different Places where the images were captured

As is seen in Figure 4.11, each participant has a different pose because he/she is allowed to act freely. Moreover, in Figure 4.11-b, different poses are shown to demonstrate varieties. Accordingly, each joint point of a subject is not clearly visible at all times, and that is, this handicap is to be overcome anyway and predict gender.

4.3.2 Data Expansion

In the previous section, "Data Collection" is explained with reasons why a new dataset should be created as well as how such data is collected. However, the data for 60 people is not large enough for age prediction for different ranges; hence, data collection continued as long as possible. Eventually, 110 more people participated as volunteers and a dataset with a total of 170 persons' data was created out of which 65 are females and 105 males between the ages of 15 to 72.

Each participant is free to take any position or do any act while his/her images are captured, as seen in Figure 4.11(a), and joint points' coordinates are taken. In addition, all is free to remain in front of the camera for as long as they wish; hence the duration spent before the camera differs for each participant.

In every second, 30 frames are captured with a 3D camera and, for each frame, the images of a scene and the coordinates of joint points are saved. Totally, 103,982 scene data are captured, 42,106 of which belong to females and 61,876 to males.

4.3.3 Classification Methods

In this part of the study, a multi-layer neural network is generated where a sigmoid function is used in the hidden layer with different numbers of neurons (10 to 50) and softmax transfer function takes place in the output layer. Additionally, the Levenberg-Marquardt training algorithm is used to optimize the weight, bias and error.

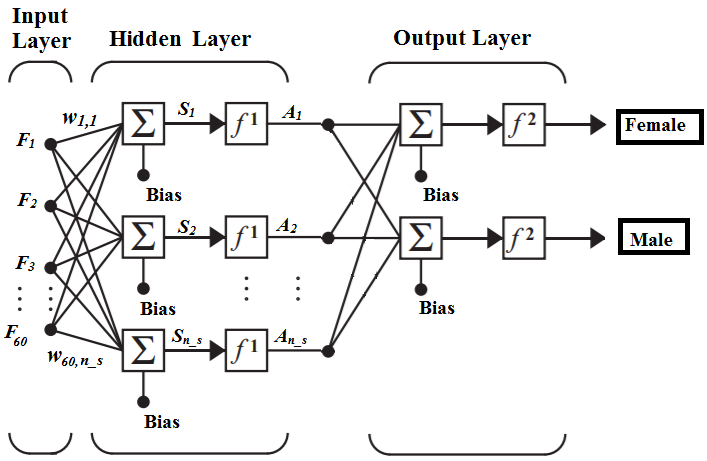


Figure 4. 12. Two-Layer Network for Gender Prediction

Figure 4.12 is the structure of our ANN with two layers; one of them is the hidden layer and the other one is the output layer. In the Input layer, there are 60 inputs (20 joint points with 3 coordinates, 20*3). In the hidden layer, the number of neurons is changed from 10 to 50, increasing by fives. The test results are shown according to these number of neurons. The output layer has 2 neurons that represent male and female.

4.3.3.1 All-Body Joint Points Features

According to the above procedure, the obtained coordinate points are used as input for the neural network. Firstly, all 103892 scenes' data are used by dividing 15% of them for validation, another 15% for testing, and the remaining 70 % as the training data. In the hidden layer, the number of neurons is changed starting from 10 to 50, increasing by fives, and the results are recorded as shown below. In Table 4.6, the number of correctly and incorrectly labeled scene data with the accuracy rates are shown for *Train*, *Test*, *Validation* and *All* as the summation of all these three groups.

Table 4. 6. Confusion Matrix Using All-Body Joints Coordinates

Confusion Matrices Using All Body Joints Coordinates												
	Train			Validation			Test			All		
Number of Neuron	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy
10	69507	3281	95,49	14913	684	95,61	14857	740	95,26	99277	4705	95,48
15	70611	2177	97,01	15113	484	96,90	15135	462	97,04	100859	3123	97,00
20	70658	2130	97,07	15157	440	97,18	15146	451	97,11	100961	3021	97,09
25	72051	737	98,99	15410	187	98,80	15416	181	98,84	102877	1105	98,94
30	**72328**	**460**	**99,37**	**15493**	**104**	**99,33**	**15473**	**124**	**99,20**	**103294**	**688**	**99,34**
35	72318	470	99,35	15486	111	99,29	15482	115	99,26	103286	696	99,33
40	72261	527	99,28	15451	146	99,06	15473	124	99,20	103185	797	99,23
45	71645	1143	98,43	15354	243	98,44	15336	261	98,33	102335	1647	98,42
50	**72371**	**417**	**99,43**	**15484**	**113**	**99,28**	**15482**	**115**	**99,26**	**103337**	**645**	**99,38**

As stated in the table, the maximum accuracy rate of all parts (Test, Validation and Test) is 99.38 % for the number of neurons in the hidden layer, which is 50. Though, it is also seen in Table 4.7 that the accuracy rates of the Test parts both for 30 and 50 neurons are the same at 99.26%. The time requirement to process the body joint points coordinates of 30 images is 0.12 seconds for the best accuracy values obtained using 50 neurons in the hidden layer of ANN. To calculate the elapse time, again 30 images are considered because 30 images are captured by the 3D camera in one second.

After examining the gender prediction accuracies for different number of neurons and finding the best result, it is important to achieve higher accuracy results for either the upper body or the lower body joints' coordinates.

4.3.3.2 Upper Body Joint Points Features

Only the first 11 of joint points (stated in Table 4.7) are used for Upper Body Parts. With the same test, validation and train percentage ratios of all the scenes, this ANN process is repeated for upper and lower body parts. The following two tables: Table 4.7 and Table 4.8 show the results of the accuracy rates for different neuron sizes in the hidden layer of ANN.

Table 4. 7. Confusion Matrix Using Upper Body Joint Coordinates

Confusion Matrices Using Upper Body Joints												
	Train			Validation			Test			All		
Number of Neuron	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy
10	63332	9456	87,01	14534	2063	87,57	13595	2002	87,16	90461	13521	87,00
15	63629	9159	87,42	13634	1963	87,41	13695	1902	87,81	90958	13024	87,47
20	65851	6937	90,47	14142	1455	90,67	14089	1508	90,33	94082	9900	90,48
25	69612	2876	96,03	15002	595	96,19	15036	561	96,40	99950	4032	96,12
30	68103	4685	93,56	14578	1019	93,47	14539	1058	93,22	97220	6762	93,50
35	68019	4769	93,45	14560	1037	93,35	14585	1012	93,51	97164	6818	93,44
40	70837	1951	97,32	15134	463	97,03	15138	459	97,06	101109	2873	97,24
45	**71259**	**1529**	**97,90**	**15287**	**310**	**98,01**	**15258**	**339**	**97,83**	**101804**	**2178**	**97,91**
50	67590	5198	92,86	14455	1142	92,68	14499	1098	92,96	96544	7438	92,85

Table 4.7 shows the accuracy rates of the ANN result for gender prediction with only the upper part of the body joint coordinates. The best accuracy for all (test, validation and train) is 97.91% when the number of neurons in the hidden layer is 45. Also, each Train (97.90%), Validation (98.01%) and Test (97.83%) accuracies are the best among different numbers of neurons in the hidden layer of networks.

4.3.3.3 Lower Body Joint Points Features

The last 10 of the joint points (as in Table 4.7) are used for the Lower body Part system inputs, and the following table, Table 4.8, presents its accuracy.

Table 4. 8. Confusion Matrix Using Lower Body Joint Coordinates

Confusion Matrices Using Lower Body Joints												
	Train			Validation			Test			All		
Number of Neuron	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy	Correct Labeled	Wrong Labeled	% Accuracy
10	67293	5495	92,45	14439	1158	92,58	14462	1135	92,72	96194	7788	92,51
15	61940	10848	85,10	13248	2349	84,94	13279	2318	85,14	88467	15515	85,08
20	67193	5595	92,31	14393	1204	92,28	14357	1240	92,05	95943	8039	92,27
25	70141	2647	96,36	14970	627	95,98	15003	594	96,19	100114	3868	96,28
30	64959	7829	89,24	13912	1685	89,20	13904	1693	89,15	92775	11207	89,22
35	64365	8423	88,43	13835	1762	88,70	13724	1873	87,99	91924	12058	88,40
40	**70915**	**1873**	**97,43**	**15184**	**413**	**97,35**	**15174**	**423**	**97,29**	**101273**	**2709**	**97,39**
45	70062	2726	96,25	15038	559	96,42	15014	583	96,26	100114	3868	96,28
50	67737	5051	93,06	14492	1105	92,92	14539	1058	93,22	96768	7214	93,06

Table 4.8 shows the results concerning the accuracy rates of different neuron sizes ranging from 10 to 50 in the hidden layer of networks, whose inputs are the Lower body parts' joint point coordinates. The best accuracy value is 97,43 % for Train, 97.35 % for Validation, 97.29 % for Test and 97.39 % for all. All these accuracy values are obtained from 40 neuron sizes in the hidden layer of the network.

Up to this point, all the scenes' data are used to train, test or validate. However, each person among the participants should be taken individually to test each network trained and validated with others. For this reason, the *Leave One Person Out* (LOPO) method is applied, and 170 distinct networks are obtained for each person, whose scene data are individualized for testing. First of all, the LOPO method is applied using all joint point coordinates as inputs. Since both 50 and 30 neuron sizes in the hidden layer have the same accuracy values, 30 neuron sizes in the hidden layer is preferred due to its simplicity compared to 50 neuron sizes in the hidden layer. Under these circumstances, Figure 4.12 shows the histogram of the accuracy rates, rounded to the nearest whole number, for 170 people.

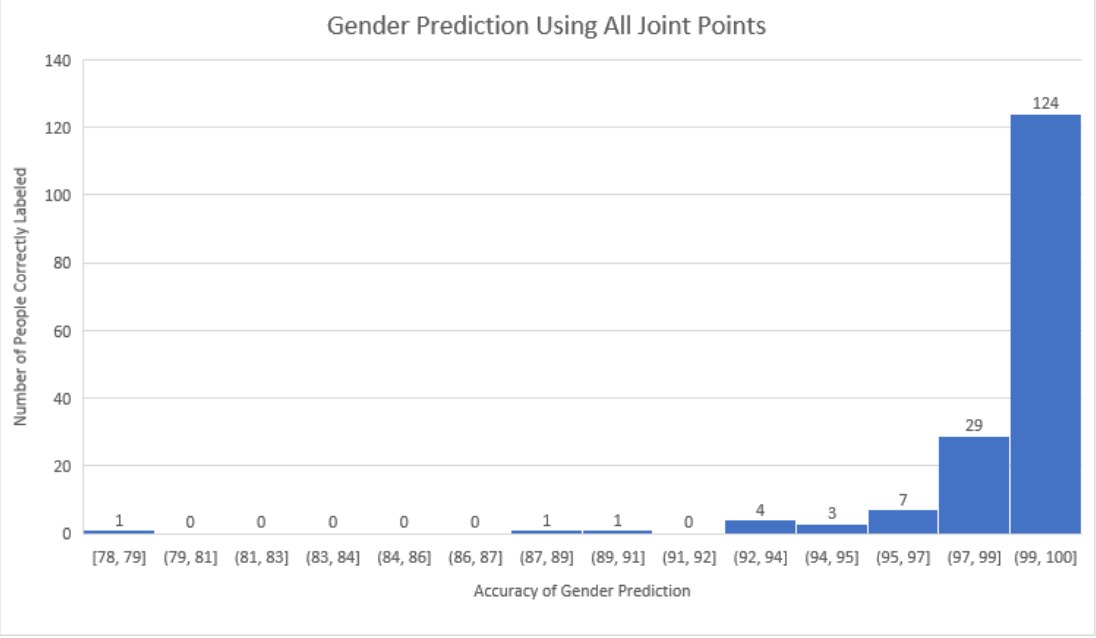


Figure 4. 13. Histogram of the LOPO method with All Joint Points of the Body

Figure 4.13 represents the accuracy rates of the human gender prediction, and the number of people corresponds to the accuracy rates' histogram. These rates define the result of ANN for each person's gender if he/she is male/female. If male and the result of ANN is 99 % male and 1 % female, it means that the ANN is successful at 99 %

accuracy. As seen in Figure 4.13, out of 170, 124 people are classified 99-100 %, 29 at 97-99 % and 14 are classified correctly between 92 and 97 per cent. Only 3 of them are classified with less than 91% accuracy rate, and they can be taken as outliers.

Then, the same procedure is repeated with the same neuron size in the hidden layer, but with different input parameters as upper and lower body parts' joint point coordinates. The following two figures show the histogram of the accuracy rates for those whose upper or lower body parts' joint point coordinates are considered to be inputs for networks with 30 neurons in the hidden layer in order to be in the same condition.

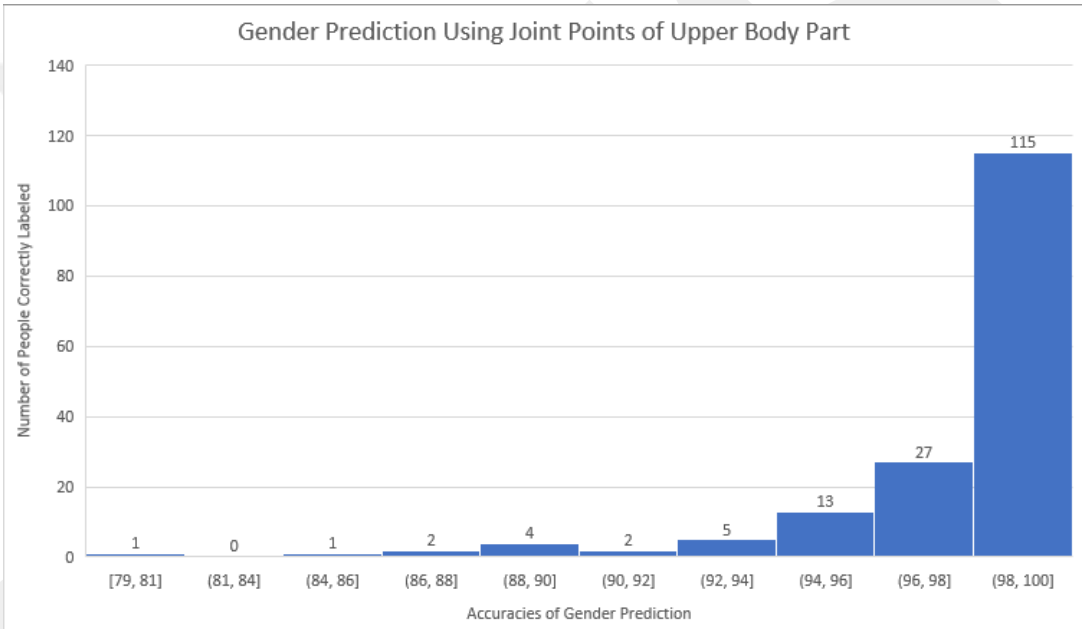


Figure 4. 14. Histogram of the LOPO Method with Upper Joint Points of Body

In this case, when the Upper body parts' joint points are examined, the number of individuals, shown in Figure 4.14, in the same interval decreases compared to the previous histogram. Nonetheless, there are still 115 volunteers correctly classified in a range of 98-100 %, and 8 classified correctly less than the range of 90%. This time, not all 8 may be the outlier, but at least 2 of them may be so, the accuracy rates of which are 79-81 % and 84-86 %.

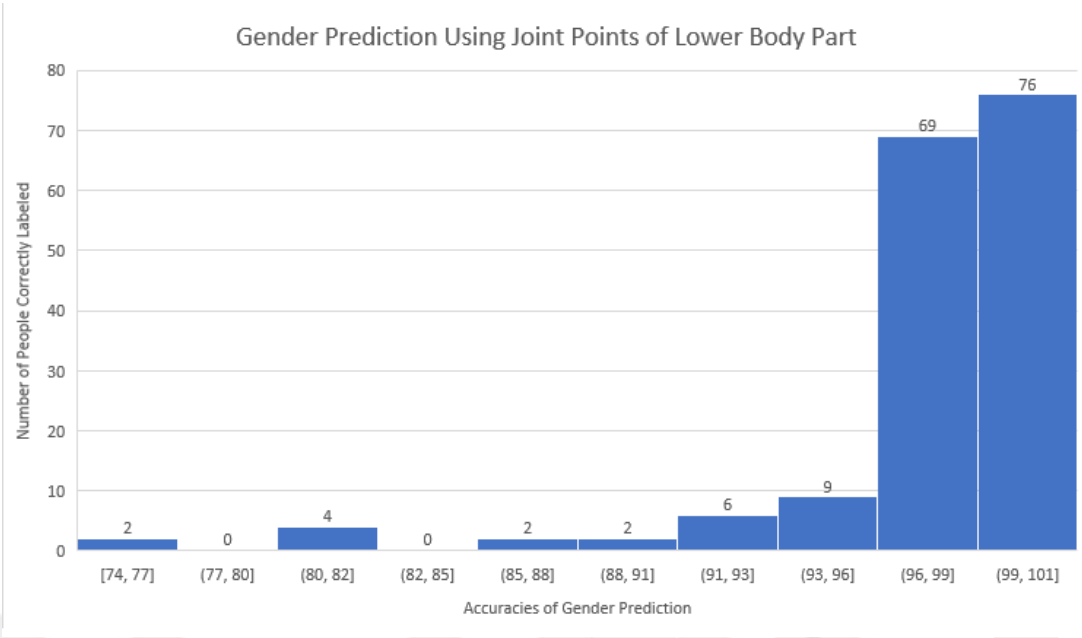


Figure 4. 15. Histogram of the LOPO method with Lower Joint Points of Body

When the Lower body parts' joint points are examined as the input, again the number of individuals in the same interval decreases compared to the previous two histograms. Although this is expected for the first histogram, it is unexpected for the second one. According to Tables 4.9 and 4.10, the best accuracy values are 97.83 % for the Upper Body Part and 97.39 % for the Lower Body Part. Nevertheless, both the Upper and Lower Body Parts have different neuron sizes in the hidden layer for whose control, the last experiment is repeated with a 40 neuron size, yielding the best accuracy rate among the others as Table 4.10 depicts. Figure 4.16 shows the histogram representation for the Lower body part joint point coordinates taken as inputs and the LOPO method applied.

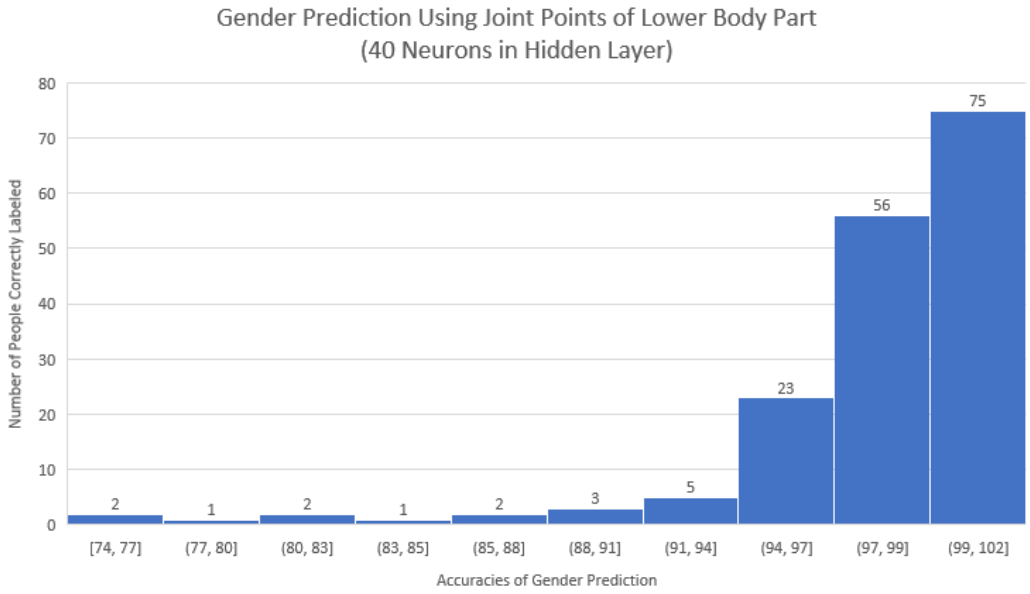


Figure 4. 16. Histogram of the LOPO method with Lower Joint Points of Body With 40 Neuron in the Hidden Layer

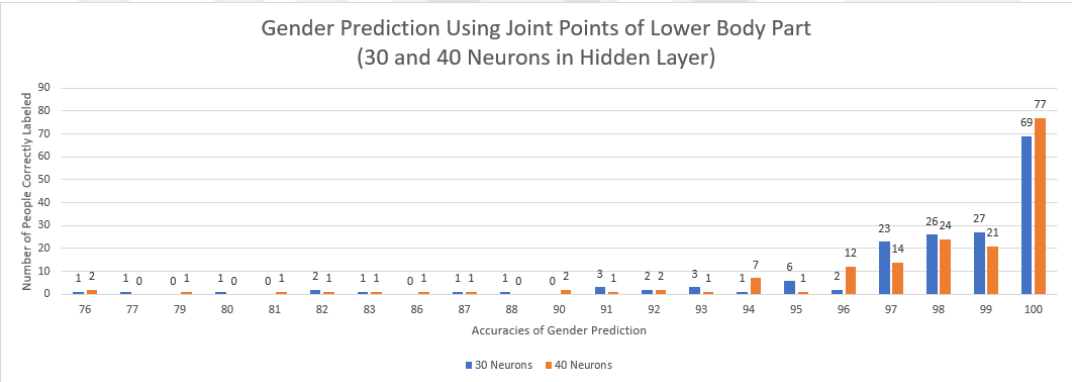


Figure 4. 17. Histogram of the LOPO method with Lower Joint Points of Body With 30 and 40 Neurons in the Hidden Layer

Figure 4.17 shows both histograms representing the accuracy values obtained from ANN with 30 and 40 neurons in the hidden layer. The blue bars symbolize the 30 neurons and orange bars symbolize the 40 neurons in the hidden layer. If it is necessary to start comparing the highest 3 accuracy values (accuracy ≥ 98%), the sum of each blue or orange bar values are the same with the value 122. However, there are differences in each accuracy value as is seen in Figure 4.17. For example, for the 30 neurons, there are 69 with 100% accuracy rate, but for the 40 neuron sizes, there are 77 people classified correctly in the same accuracy rate. On the other hand, in accuracy values of 98% and 99%, 30 neuron sizes provide better classification than 40 neurons. As a result, the number of people does not vary in the highest accuracy values for the neuron size in the hidden layer.

4.3.4 Gender Prediction from 3D Body Data Using ANN - Conclusion

In this study, the proposed approach is a new perspective to predict gender based on 3D postures of the human body. 3D skeleton data, which consists of the 3D coordinates of joint points, is used as ANN input. The used classifier is a Multilayer Neural Network and the number of layers in the hidden layer is varied in the experiment to test and increase the accuracy values. In order to verify the classifier, the input data is divided into three groups: train data, validation data and test data. The data used in the classifier is collected from 170 volunteers: 65 females and 105 males between 15 to 72 years of age. During data collection, a depth camera is used and volunteers are free to take position in front of the camera and strike any pose. By using 3D skeleton data, the accuracy results are represented in three parts: all body 99.38%, upper body 97.91%, and lower body 97.43%. These results are obtained whit a multilayer neural network with 30, 45, and 40 neuron sizes in the hidden layer, respectively.

In this posture-based system, using the 3D skeleton data with the help of a 3D camera, gender can be predicted automatically in 0.12 seconds, which is nearly real-time. The main contribution of this study is that even if the camera cannot capture the whole-body parts, the system obtains an accurate result for gender prediction using the upper or lower part of the human body. In crowded places, capturing high-resolution face images and whole joint points of the body is difficult. Yet, our approach only requires to capture joint points of the body parts for gender prediction.

Comparing the proposed method with other approaches is difficult due to the use of different databases by different researchers. Still, our method is proven to be more successful using 3D data captured from a significant number of real people. The previous studies, as per the 'Related Study' part of this work focusing on the Kinect camera, use databases with small sample sizes and lower accuracy values than ours. The exception is Kakadiaris et al. [15], who proposed a study using upper and lower body parts of the body using a Kinect camera, but the accuracy values (86.0% all body, 78.0% lower body, and 72.0% upper body for real images) are not better than ours. The Kakadiaris method is more powerful in the CAESAR database, which offers data different from that of Kinect. Tran et al. [143] proposed a study only using upper parts of the body, again testing with a small sample group.

CHAPTER 5

AGE PREDICTION

Automatic age prediction is one of the main incentives behind this study. There are two approaches to predict the age of persons: face-based and body-based approach. In the first part, the experimental results of the face-based approach are explained. In the second part, the results of the body-based approach are presented. In the face-based approach, there are also two methods; one of them is the statistical pattern recognition method and the second one is deep learning. The results of the first approach involving statistical pattern recognition approach are given in Section 5.1, deep learning findings are presented in Section 5.2, and the results of body-based artificial neural network are given in Section 5.3

5.1 Age Prediction from Face Images Using Statistical Methods

In this part of the study, LBP and HOG features are used for feature extraction, and the theoretical background about these methods have been described in Section 3.2. Additionally, the KNN and multi-SVM classification methods are used for classifying the age of individuals as described in section 3.3.

5.1.1 Face Image Database for Age Prediction using Statistical Methods

In this part of the study, the Gallagher's database [87] is used to predict age from face images. This database contains 28231 faces and 5080 images in different ages between 0 and 75. This database contains 28231 faces and 5080 images in different ages between 0 and 75. The images are from weddings, and family and group images and faces from these images are labeled with age and gender, enabling one to use them for age and gender prediction. There are 8959 gray-scale face images taken without any environmental control at dimensions of 49x61 pixels. Unfortunately, the images are not equally distributed over the age ranges in this database. Table 5.1 shows the age

range and the number of the images in each age range. Some example images from this dataset are given in Figure 5.1.

Table 5. 1. Age Ranges and the number of images with the corresponding Gallagher's database.

Classes	Age Range	Number of Face Images
1	0 - 2	493
2	3 - 7	789
3	8 - 12	519
4	13 – 19	682
5	20 – 36	3,539
6	37 – 65	2,235
7	66+	702

Figure 5. 1. Example images in the Gallagher's Database

5.1.2 Age Prediction with LBP Feature Extraction Method

The local binary pattern (LBP) feature extraction method is one of the common methods using texture information in the area with small changes like face texture. Therefore, for age prediction, this method is used to extract features, and the experimental results with two classification methods (namely SVM and KNN) are shown below. In order to test the accuracy, the database is divided into two parts: training (70% of the database) and testing (30% of the database).

Table 5. 2. Accuracy Rates of LBP with SVM classification method

LBP - Multiclass SVM		
	One-versus-One	**One-versus-All**
Linear	**40.1%**	33.2%
RBF	36.9%	26.6%
Polynomial	21.9%	21.1%

In Table 5.2, the classification results are shown with SVM classifier using LBP features. The LBP feature extraction method has been experimentally tested with two multi-class SVM classification methods; one-versus-one and one-versus-all with three kernel functions: Linear, RBF, and Polynomial. As seen in the table, the linear kernel function with one-versus-one multi-class SVM method gives the best accuracy rate at 40.1%. Additionally, it is clear that the one-versus-one SVM methods' accuracies are greater than one-versus-all SVM methods accuracies for all three kernel functions. Because the SVM classification method is a binary classifier, considering the pair of two classes in a multi-class SVM classification is more powerful than the one-versus-all method [144].

In Table 5.3, the confusion matrix of one-versus-one multi-class SVM classification method with linear kernel function is shown, and it is clearly seen that most of the images are labeled as fifth class whose age range is between 20 and 36. The number of images of the fifth class is 3,539, which is the highest number in the database. Therefore, most of the images are classified as fifth class in the trained system. To classify 30 images in the Gallagher's database, the elapse time is 271.71 seconds using LBP + SVM methods.

Table 5. 3. Confusion matrix of age prediction with One-versus-one SVM classification method for LBP.

One-versus-one Multiclass SVM with Linear Kernel Function							
Classes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	20	26	8	4	1	1	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	128	130	197	1058	236	669	211
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	43
Accuracy (%)	**13.5**	**0**	**0**	**0**	**99.6**	**0**	**0**
Total Accuracy = 40.1%							

The features obtained by applying the LBP method on face images are classified using the KNN classification method, and the accuracy rates based on the k values are tabulated in Table 5.4.

Table 5. 4. Accuracy Rates of KNN classification method for LBP.

	KNN (30% test and 70% train of database)												
k	***1***	***3***	***5***	***7***	***9***	***11***	***13***	***15***	***17***	***19***	***45***	***81***	***100***
Accuracy (%)	19.9	27.0	29.8	30.9	32.2	32.8	32.9	33.4	34.6	35.3	37.6	**38.4**	37.7

As seen in Table 5.4, the most accurate k value in the KNN classification method is k=81 with 38.4% accuracy. The first accuracy rate is 19.9 % for k=1 and with increasing k values, accuracy also picks up to k=81. At first sight, 81 for the k values may be high; yet, the minimum number of images in a class of the database is 493, which is large enough for the k=81 value. However, the accuracy rate increasing from k=45 to k=81 is not as high as should be expected. Table 5.5 shows the confusion matrix of LBP and KNN classification methods with k=81. The time required to classify the people's age in 30 images from the Gallagher's database is 2.23 seconds using both LBP and KNN methods when k=81.

As in Table 5.3, the most classified class is the fifth class for the KNN classifier, as the same seen in Table 5.5. Since the highest number of image falls into this class, most images are classified as such.

Table 5. 5 Confusion matrix of age prediction with KNN classification method for LBP

	KNN with k=81							
		Actual						
	Classes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Predicted	**1**	8	15	4	4	0	0	0
	2	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0
	5	129	125	191	969	223	615	204
	6	9	16	10	88	14	55	7
	7	2	0	0	1	0	0	0
	Accuracy (%)	**5.4**	**0**	**0**	**0**	**91.2**	**8.2**	**0**
	Total Accuracy = 38.4%							

5.1.3 Age Prediction with the HOG Feature Extraction Method

The Histogram Oriented Gradient feature extraction method (or HOG for short) is also a successful method for face images to predict age because it uses a texture feature by counting the occurrences of gradient orientation in the localized portions of an image. Additionally, for age prediction, HOG feature extraction methods are experimentally tested with the SVM classifier with the accuracy results shown in Table 5.6.

Table 5. 6. Accuracy rates of HOG feature extraction with the SVM classification method using Linear, RBF and Polynomial kernel functions.

HOG - Multiclass SVM		
	One-versus-One	**One-versus-All**
Linear	**36.0%**	**36.0%**
RBF	31.3%	26.9%
Polynomial	20.0%	19.9%

Table 5.6 depicts the accuracy rates of the HOG method for age prediction according to the SVM classifier. Three kernel functions - linear, RBF and polynomial - are tested using one-versus-one and one-versus-all multi-class SVM classification. As can be clearly seen in the table, to predict the age factor, the highest accuracy results of the classification with SVM classifier using HOG features belongs to the Linear kernel function in both multi-class methods. For the best accuracy of these methods, the required time to classify 30 images is obtained as 506.74 seconds. On the other hand, polynomial functions do not result in satisfactory outcomes, which means that HOG-assisted age prediction may not be solved by polynomial functions.

Not only the SVM classification methods, but also the KNN classification method is applied throughout age prediction experiments within this field of study. In this respect, the KNN classification method is applied with different *k* values, for which Table 5.7 shows the accuracies of age prediction results obtained by the KNN classification method.

Table 5. 7. Accuracy Rates of KNN classification method for HOG.

	KNN (30% test and 70% train of database)												
k	***1***	***3***	***5***	***7***	***9***	***11***	***13***	***15***	***17***	***19***	***45***	***87***	***100***
Accuracy (%)	19.9	26.5	30.3	31.8	33.4	34.4	34.9	35.0	35.0	35.2	35.3	**36.6**	36.5

It can be noted in Table 5.7 that the accuracy values tend to elevate with increasing k values up to k=87. Although for the first k value, the accuracy is 19.9% - a percentage which is quite low - increasing the k value after k=13 does not affect the accuracy any more than 10 percent. Nevertheless, the most accurate rate is 36.6% for k=87 – still not as high as expected due to the large size of the dataset for each class of the age range. Given that the best accuracy is obtained at k=87, the elapsed time for the classification process is calculated at this value and found to be 7.32 seconds. The confusion matrix of the KNN classification method with k=87 appears in Table 5.8., based on which one can see that the matrix mostly concentrates on two classes: fifth and sixth which, in turn, have the highest number of images - 3,539 and 2,235 images, respectively; therefore, the most successful classification results go to the fifth class with the highest number of images.

Table 5. 8. Confusion matrix of age prediction with the KNN classification method for HOG

KNN with k=87								
		Actual						
	Classes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Predicted	**1**	26	32	10	1	0	0	0
	2	11	12	4	3	0	0	1
	3	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	0
	5	68	78	128	816	172	528	152
	6	43	33	63	240	63	141	58
	7	0	1	0	1	2	1	0
	Accuracy (%)	**17.6**	**0**	**0**	**0**	**76.8**	**21.0**	**0**
Total Accuracy = 36.6%								

5.1.4. Comparison of Feature Extraction Methods

For age prediction from face images, two feature extraction methods - LBP and HOG - are used on those appearing on the Gallagher's dataset, with the SVM and KNN results shown together in Table 5.9.

Table 5. 9. The highest accuracy rates of LBP and HGO feature extraction, SVM and KNN classification methods and elapsed time to process 30 images

30% Test 70% Train	**LBP**		**HOG**	
	Accuracy (%)	**Time (s)**	**Accuracy (%)**	**Time (s)**
SVM	40.1 %	271.71 s	36.0 %	506.74 s
KNN	38.4 %	2.23 s	36.6 %	7.32 s

As seen in Table 5.9, the most accurate feature extraction and classification method combination is LBP+SVM with 40.1 % accuracy, whereas the KNN accuracies are less than SVM classification with the LBP extraction method. On the other hand, in case of the HOG method, the KNN accuracy value is greater than SVM, though still standing at a very close range. The elapsed time to process 30 images of LBP + KNN

is the least, and HOG + SVM is the most. It is seen in Table 5.9 that the KNN time requirement is far less than SVM's in both LBP and HOG. Also, both for SVM and KNN, the HOG method requires more time to process, whereas LBP requires less.

5.1.5 Age Prediction Using Statistical Methods - Conclusion

The study of age prediction is explained step by step in detail. The LBP and HOG are used for feature extraction and SVM and KNN are applied for classification. At this stage, the results of the experiments are presented as per the proposed methods.

Here, face images are used to predict the people's age according to the archives of the Gallagher's database. Previous studies [87]–[89], [92], having also utilized this database, attempted to predict age and gender. In age prediction, the performance of the related studies [87]–[89] are 42.9%, 63.01%, 66.6%, respectively, having incorporated more than one feature extraction method (such as FPLBP [89], SIFT and CH [88] on top of LBP) and involving more pre-processing steps (such as PS and FPS) for illumination reduction in [92] and dimension alignment. Therefore, there can be slight differences in the performances between the previous studies and the present work. In comparison, the performance of Khalifa's [14] method, applied on the FG-Net dataset, is found to be significantly higher than ours although both have used the same feature extraction and classification methods. The reason for this is the use of two different databases. Whereas the Gallagher's database has gray-scale images at 49 x 61 - much smaller in comparison, the FG-NET database offers 640 x 480 images with RGB and gray-scale in addition to a high resolution. What's more, our tests are realized upon 30% of all the dataset and 70% train data of all the dataset, while other studies use either Leave-One-Out or 5-fold/10-fold cross validation techniques to determine the classifier's performance. In all, fewer images are used in the present work for training as opposed to others. The reason for selecting our cross-validation technique in this way is to enable a comparison between the CNN technique and statistical techniques under the same conditions and the same databases.

5.2 Age Prediction from Face Images Using Deep Learning

As stated in section 3.5, CNN's are recently being used in different fields of computer vision one of which is age prediction from face images. For this purpose, to compare the statistical methods (SVM, KNN) with CNN, this part of the study shows the CNN experimental results for age prediction from face images. To do so, transfer learning is used as deep learning by applying a CNN model, which is pre-trained for a specific purpose. In transfer learning, this pre-trained model is used for another purpose to save resources and time, clearly reducing the training time in effect. Figure 5.2 represents the transfer learning allowing for the re-use of the pre-trained networks.

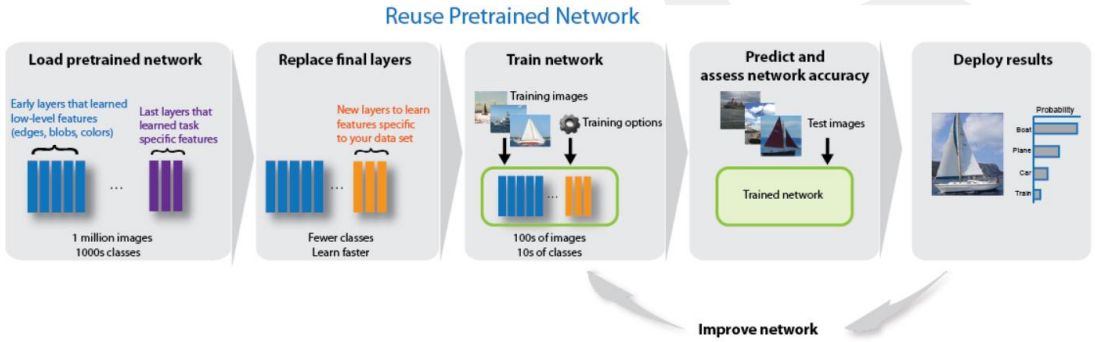


Figure 5. 2. Transfer Learning Representation[145]

As seen in Figure 5.2, there are five main parts within transfer learning; loading the pre-trained network, replacing the final layers, training the network, predicting and assessing network accuracy, and deploying the results. At first, pre-trained network is loaded in order to use pre-trained model with its layers. In this network, leading layers are used to learn low-level features; such as edges, blobs, and colors; while the last layers are used to learn task-specific features. Since being task-specific, the last three layers of the pre-trained network are eliminated and the new layers, which are used to learn features-specific tasks in the dataset, are replaced with the removed ones in transfer learning. The next step is training the re-arranged network, whose last layers are changed. In the training step, the training images and output classes are different from the pre-trained network. Afterward, the testing step is completed with the related test images supplied to the trained network in order to predict and assess its accuracy. The last step is to deploy the results and make the network usable for the specific purpose.

In the present study, this structure is used by applying the VGG16 [122] pre-trained network, which is trained on the ImageNet dataset [146] with 41 layers and 1000 output classes. The experimental results appear in the following sections.

5.2.1 Transfer learning with IMDb and WIKI Datasets

In 2015, the competition ChaLearn LAP 2015 [147] was held on the topic of predicting age from the appearance, and the 1st winner of the competition was DEX [78], which proposed a CNN architecture to solve the appearance-based age estimation. The same structure is used in this study, that is the VGG16 architecture pre-trained on the ImageNet dataset [146] for image classification. The input size of the images, which are used for transfer learning, and the output classes are modified changed to be used in the pre-trained network as is seen in Figure 5.2. Not only the structure of the study, but also the databases [78], which were constructed using Internet images on the Wikipedia and IMDb sites, are tested to train the transfer learning of our study.

The Wikipedia and IMDb images database are the largest datasets on the internet and used to predict biological age as proposed by Rothe et al. [148]. Therefore, to predict the age of individuals, we intended to use the same datasets as published on the Website [149] with all the images and their accompanying metadata. There are two types of image datasets: full image and face-only. IMDb has 460,723 images and WIKI 62,328; we used the face-only datasets of these two websites. In the study proposed by Rothe et al. [148], it is claimed that the images age range is between 0 to 100 and a 101 classes can be classified. However, there is not an equal number of images for each age label. Additionally, the image sizes are different from each other and vary between 1 KB to 78 KB.

The first experiment is carried out with 8 classes between 0 and 80 age ranges (0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80) with different number of images in each class. For each trial, 50, 75, 80 and 90 images are taken from each group, respectively. In each experiment, the total number of images is divided into two groups: 70% for training and 30% for validation, and the results of each experiment are shown in the following figures with training and validation graphics.

According to Figure 5.3, there are two graphics which represent “Accuracy” and “Loss” of the training and validation of the network during the training phase. The blue and red lines denote the training accuracy, while the black dash lines denote the

validation accuracy. The “Results” part of the graph shows the validation accuracy in the first line, and for this training, 25% validation accuracy is achieved. Next, the “training finished” title shows the training status and whether it converges to a stable training status or not. Although the training status appears in the graphic, the percentage of the training is not shown with numerical values but, rather, with a blue line as seen in Figure 5.3, standing at nearly 95%. On this basis, the training accuracy obviously exceeds that of the test accuracy, which means that the system overfits in the training phase and learns all the training samples. However, there are samples whose correct results cannot be determined by the system. This shows that the system cannot generalize the learning. Additionally, the training start time and the elapsed time are shown about the training time in Figure 5.3.

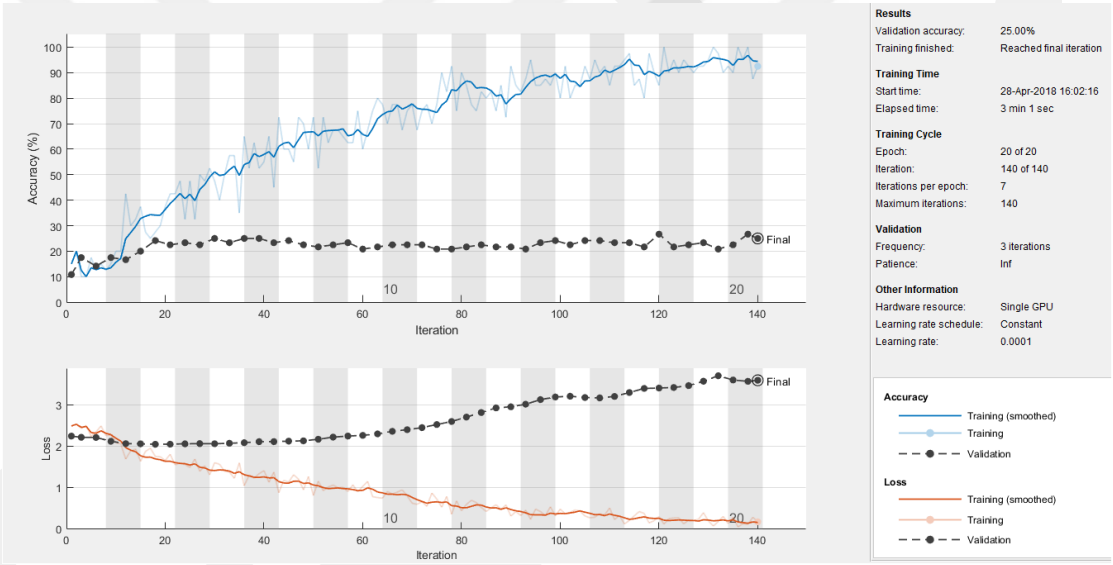


Figure 5. 3. Training Process with 8 Classes, 50 Images for each class and 20 Epochs

One of the important parts of the information board is the training cycle, which shows the epoch number, iteration, iteration per epoch, and maximum iterations.

Epoch is the number of value that passes at one time through all the training vectors to update the weights. It is represented in the figure in the background with shaded marks. The *Iteration per epoch* is rounds of updates to complete an epoch for all the training vectors. Consequently, the *maximum iteration* number is the total number of iterations for all epochs. *Iteration* shows the number of iterations during the training process. Therefore, the number of maximum epochs can be determined by the user, but the number of iteration per epoch is determined during the training process according to

the *mini batch* size, which is also assigned by the user and refers to the amount of data which is the subset of the dataset to calculate the loss and update the weights in each iteration. It has to be noted that these parameters cannot be determined and assigned under a standard set of rules because, when the problem and/or the dataset changes, the weight updates in the training process change, too.

The part related to validation provides information about the frequency of the validation, which is the number of iteration frequency for network validation, and patience of validation which is the stopping criteria of the network. Additionally, other information parts show the "hardware resources" under the titles the number of GPU, "learning rate" and "learning rate schedule".

For the 50 images in each class, the validation of the training network does not produce satisfactory results; hence, the training phase is repeated with 75 images, yielding the results as in Figure 5.4.

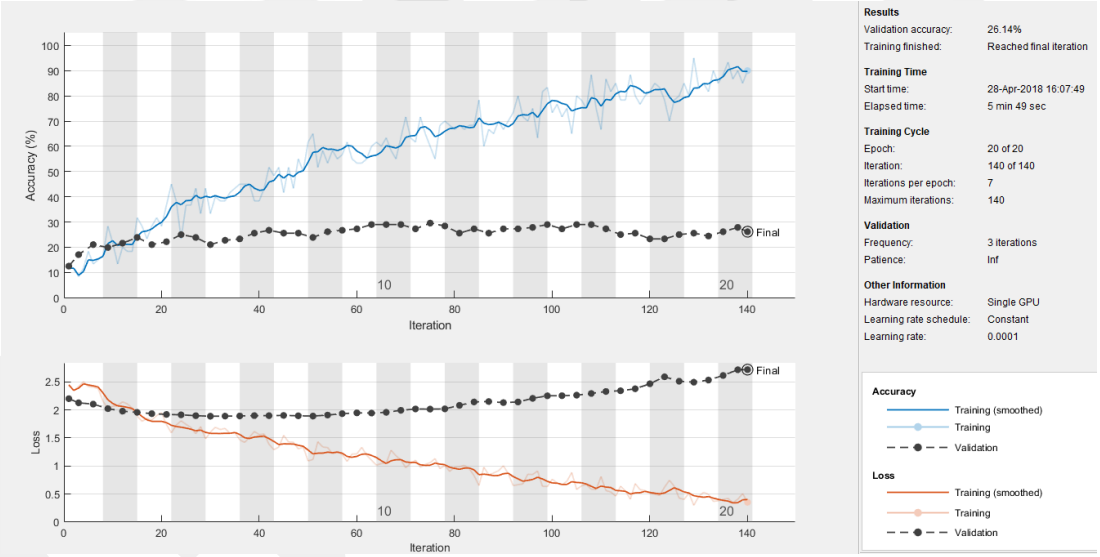


Figure 5. 4. The training Process with 8 Classes, 75 Images for each class and 20 Epochs

When the number of images in each class is increased to 75 with the same parameters - the number of epochs as 20 and mini batch size of 10 - the validation accuracy increases to 26.14%, still not satisfactory. In the next experiment, the number of images for each class is increased to 80 and the results are presented in Figure 5.5.

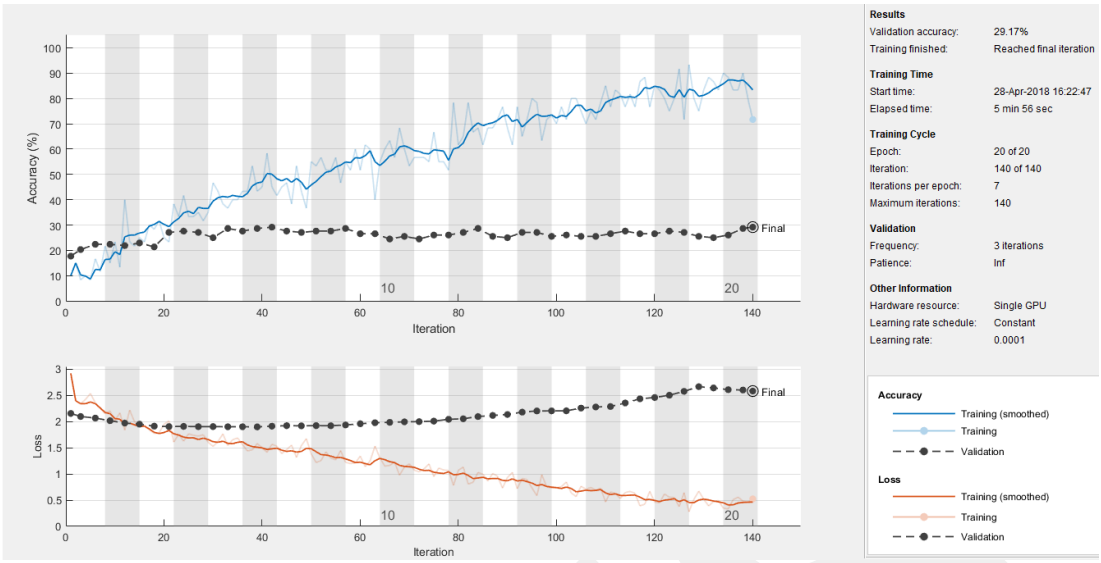


Figure 5. 5. The training Process with 8 Classes, 80 Images for each class and 20 Epochs

This time, the validation accuracy increases 29.17%, which is still very low when compared with previous studies, namely Rothe [148]. However, the number of images for the class is less in our pre-trained network.

As seen in Figures 5.3, 5.4, and 5.5, the training lines are increasingly approaching 100%, but do not become fixed after a certain point. For this reason, the same experiment is repeated, this time increasing the epoch number from 20 to 40 so that it can be stabilized after a certain point.

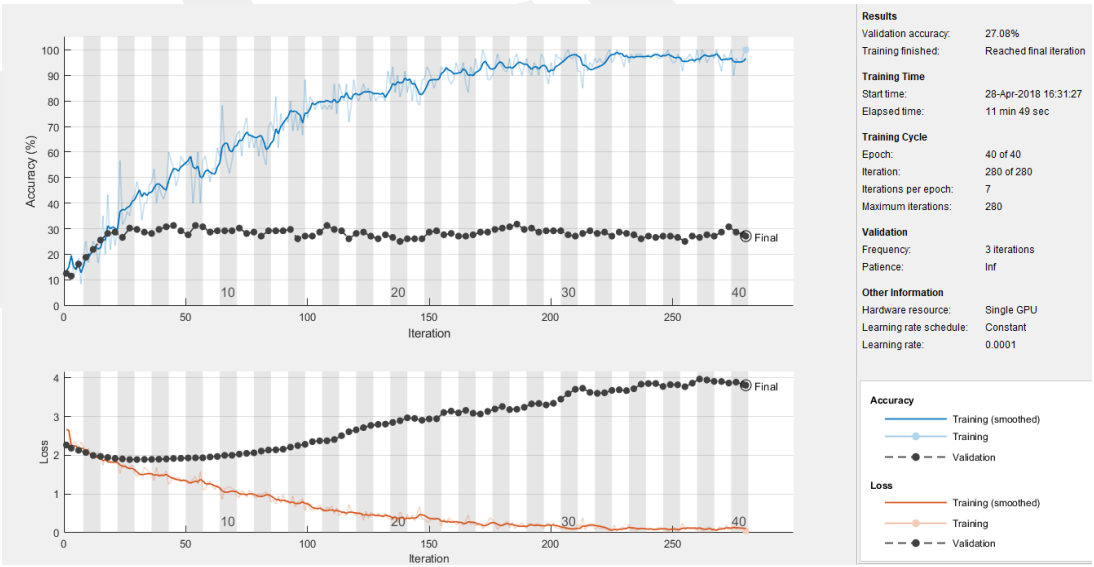


Figure 5. 6. The training Process with 8 Classes, 80 Images for each class and 40 Epochs

In Figure 5.6, the training lines of "Accuracy" and "Loss" graphics are fixed after 235 nearly iterations. However, the validation accuracy does not change after 28 iterations, and there are only slight oscillations at around 30%.

Finally, the same experiment is repeated for 90 images for each class with 60 epochs. The training and test accuracies are presented in Figure 5.7.

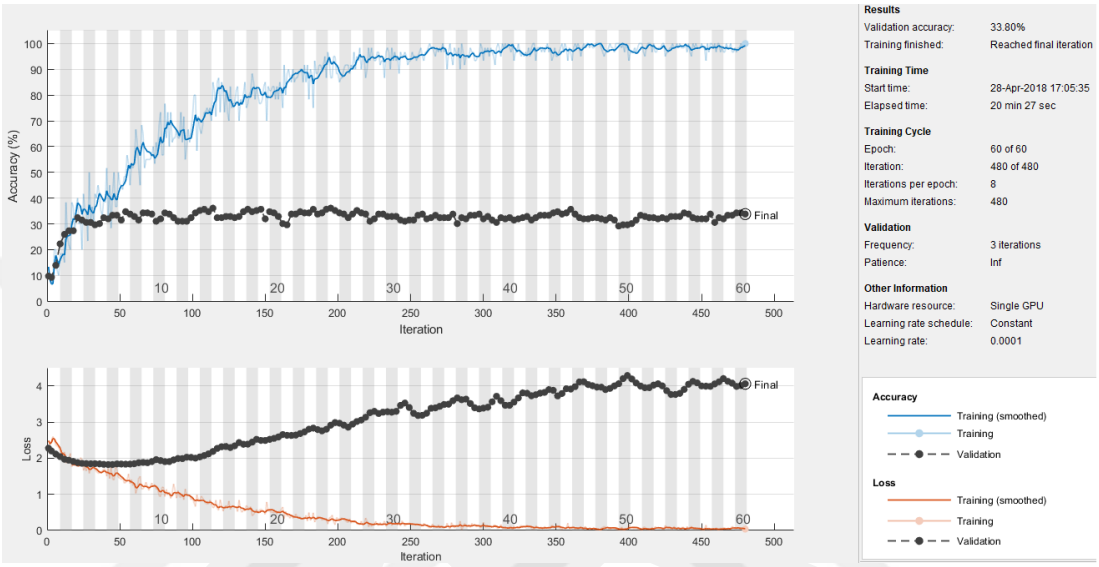


Figure 5. 7. The training Process with 8 Classes, 90 Images for each class and 60 Epochs

The behavioral patterns of the accuracy and validation lines are similar to those in the previous graphic. In the last graphic, the training accuracy line is fixed at 100% after 400 iterations, and the training loss line is fixed nearly after 380 iterations and the validation accuracy is reached to 33.80%. Although the accuracy has reached 100% and loss dropped to 0% during the training phase, the validation lines are still not as successful, as seen in Figure 5.7.

The reasons and conclusions for the obtained results can be listed as follows:

- The number of images in each class should be increased if possible. Due to shortages in the system resources, no more images can be used.
  - The hardware features of the computer used could not process any more images, being of the following properties:
    - Processor: Intel® Core™ i7-8700 K CPU @3.70 GHz
    - Installed Memory (RAM): 16.0GB
    - System Type: 64-bit Operating System
    - NVIDIA GeForce GTX 1080 with 2560 CUDA Cores

- The image sizes are not stable, and larger image sizes consume more RAM
  - As stated before, the image sizes change between 1 KB to 78 KB. If all sizes were 1KB, 77 more images would be used instead of only one image with size 78 KB. The average size of the images has not been checked; hence, the exact number is unclear.

➢ Image labelling should be more reliable than it currently is.

- It is mentioned in Rothe's study [148] that the ages of the people in the images are labeled according to both appearance (with limited annotation) and calculation of the difference between the date of birth and when the photo was taken. After the images are saved according to their calculated ages, it appears that there is still a problem with age labeling. Figure 5.8 shows example images labeled as 2-year-olds. However, as it can be seen, there are only two children with different ages and eight people more than 2 years of age. Image labelling according to such estimations between the day of birth and date of the photo may cause errors during the training of the network while updating the weights and calculating error rates. If a higher number of images is made accessible, this problem might be eradicated because of CNN's learning power.

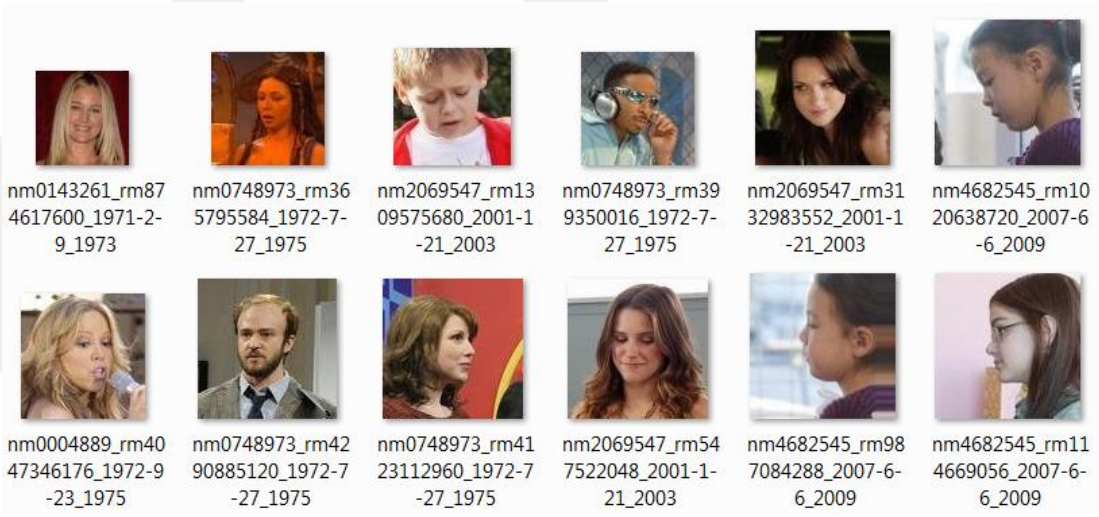


Figure 5. 8. A sample content of the folder from WIKI-IMDb Datasets with images labeled at age 2.

Due to the problems listed above, we changed the dataset used. Another face dataset proposed by Gallagher [87] and used for age and gender prediction is "*The Images of Groups Dataset*", which is used in our experiments to continue, as in the following section.

5.2.2 Transfer learning with Gallagher's Dataset

The Gallagher's dataset is created to predict the age and gender of each person among group images [87]. There are seven age groups; 1(0-2), 5(3-7), 10(8-12), 16(13-19), 28(20-36), 51(37-65), and 75(+66) and 8960 gray-scale images in total [150]. When we compare this dataset with the FGNET and WIKI-IMDb datasets, this dataset has more images than FGNET, and it is more advantageous to have image sizes of 1-2KB even though there are far fewer images and far fewer classes than WIKI-IMDb datasets.

As it is seen in Figure 5.9, there are 10 children with ages in the images as close to each other. Figure 5.9 represents a picture of the children of the average age 5 and it is much more reliable than the labeled images shown in Figure 5.8.

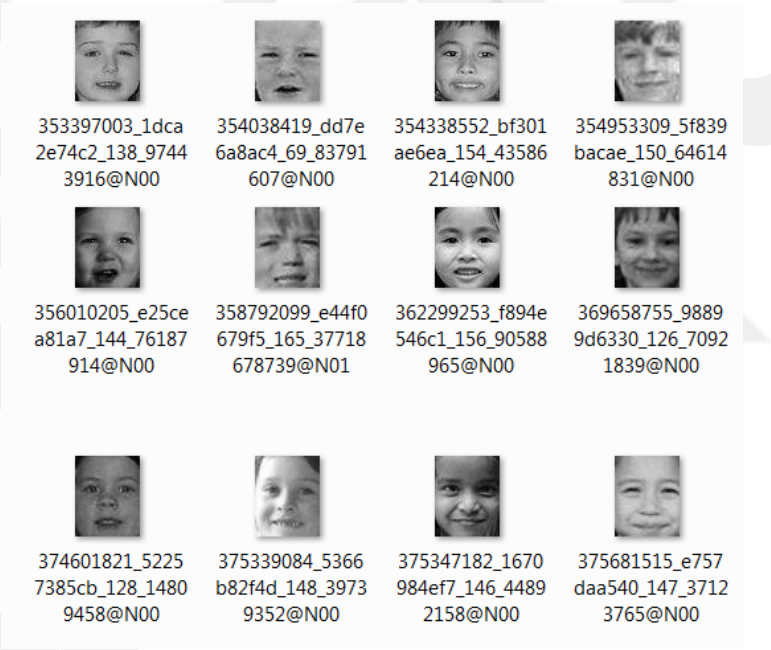


Figure 5. 9. A sample content of the folder from The Images of Group Dataset, where the images of people are labeled at age 5.

For the experiments, all the 8960 images in the dataset are used in the CNN training phase. The dataset is, then, divided into two groups: training, which is 70% of each age class in all the dataset, and validation, which is 30% of each age range in all the dataset. Transfer learning is applied for 7 classes with 10 epochs, and the results are presented in Figure 5.10.

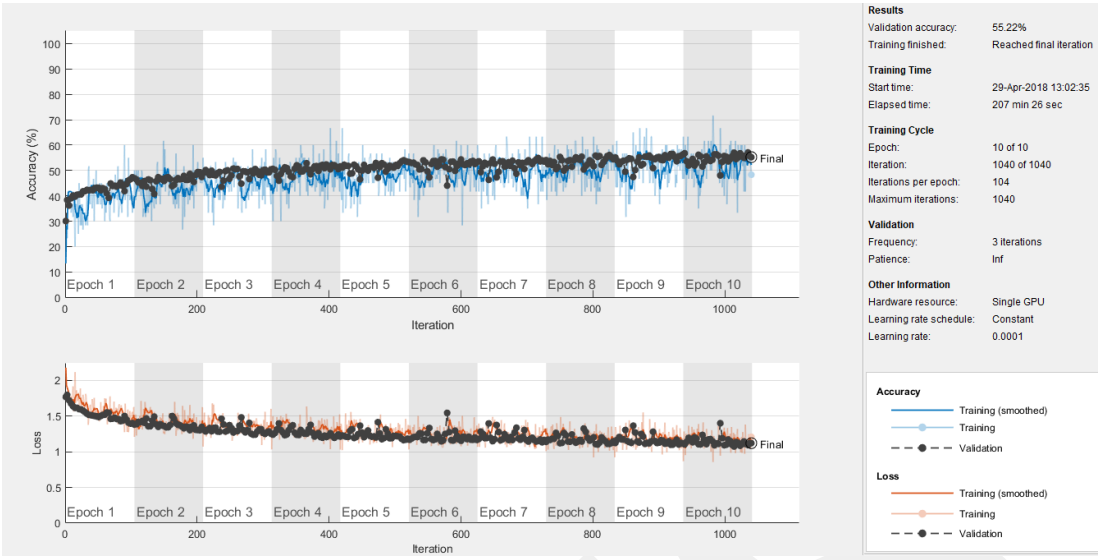


Figure 5. 10. The training Process with 7 Classes, all dataset images, and 10 Epochs

As is seen in Figure 5.10, the training and validation accuracy lines behave similarly to the training and validation loss lines. Up to 10 epochs, the validation accuracy reaches 55.22% while the lines are still not fixed; the system is retrained with more number of epochs, and the results appear as in Figure 5.11.

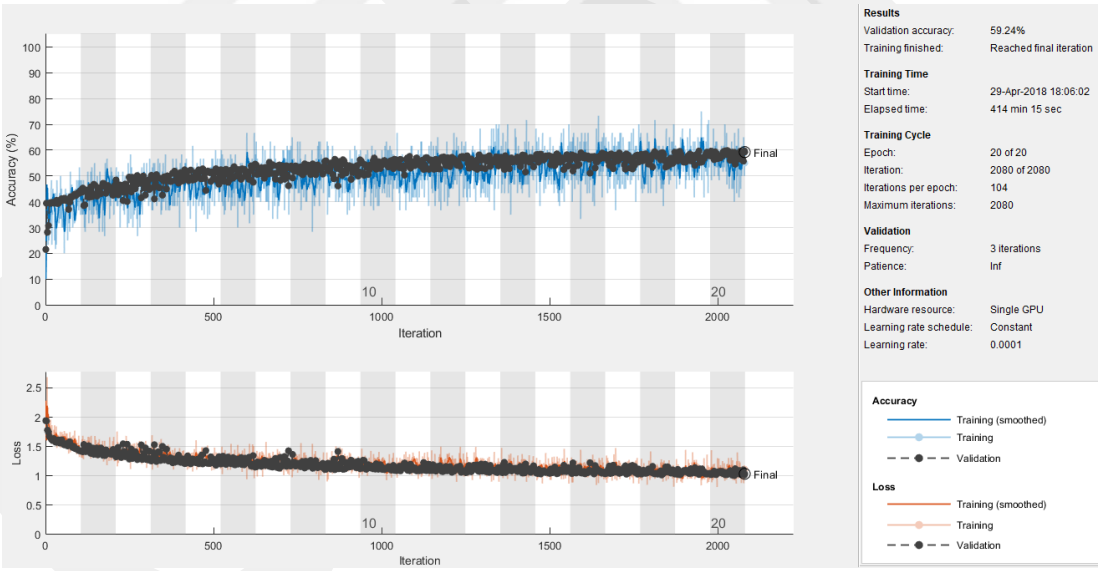


Figure 5. 11. The training Process with 7 Classes, all dataset images, 30 Epochs

When the number of epochs within the training increases, it can be seen that the validation accuracy also increases to 59.24%, as shown in Figure 5.11, thereby revealing that the number of epochs should be increased and throughout the duration the training lines are to be fixed as a straight line.

During a longer period of time as training might require, 100 epochs are assigned to the training parameters, providing results as seen in Figure 5.12.

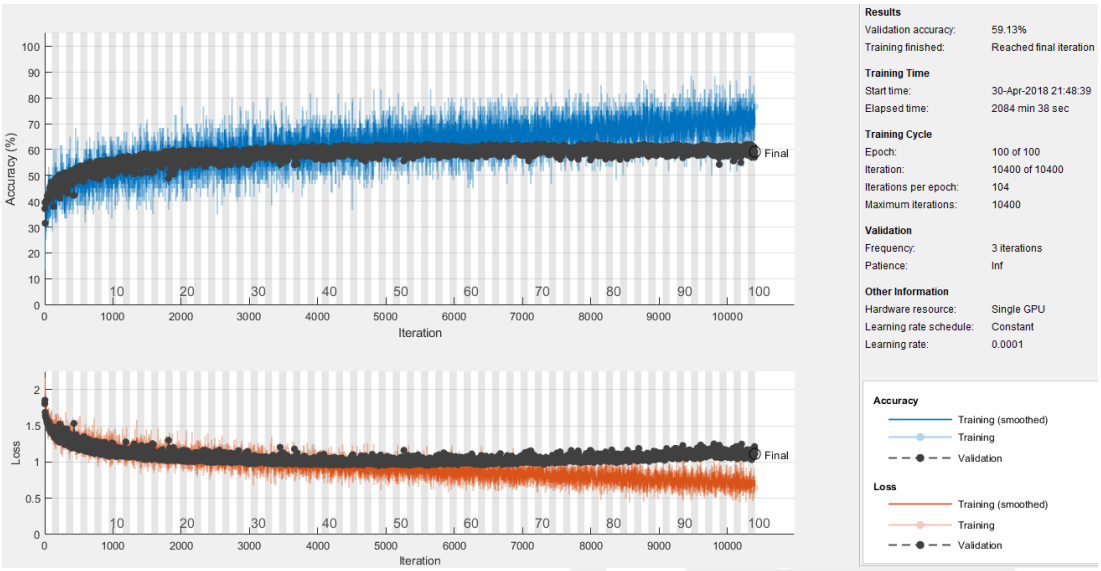


Figure 5. 12. The training Process with 7 Classes, all dataset images, 100 Epochs

In the final experiment, the training accuracy line achieves 80%, but the validation accuracy line remains unchanged up to 50 epochs. Similarly, the validation loss line does not change positively after 30 epochs, whereas the training loss line shortens and approaches 0.5. This means that the validation accuracy remains the same, and that there will be no more changes in the number of epochs as it reaches the overfitting threshold. Consequently, the final validation accuracy is 59.13%, which is less than Rothe et al.'s method with an error rate of 0.265 [148], and more than Gallagher and Chen's proposed method with an accuracy of 49.1% [87]. Under these conditions, the system is trained to obtain the best accuracy of 59.13 % and time requirement to process 30 images at 16.63 seconds.

5.2.3 Age Prediction from Face Images - Conclusion

In the present study, age prediction from face images is investigated with both statistical methods and the convolutional neural network (CNN) method on Gallagher's database. In each experiment, 30% of the dataset is used for testing and 70% for training. As mentioned in Section 5.1.5, using statistical methods such as LBP, HOG, SVM and KNN, one cannot classify low resolution images - as in the Gallagher's databased - with high accuracy. On top of this, face images, such as the ones obtained from the 3D camera used for the present study, are low-resolution images with small sizes. For this reason, training the system with low resolution images as in the Gallagher's database is important to provide accurate results in the proposed software tool. In a similar way, In a similar way, Rothe's study [148], , which

uses Wiki and IMDb database, exceeds our CNN results. However, the image size and mislabeled images complicate working on the same dataset. Thus, for the CNN method the Gallagher's database is used to predict age from face images. The best results of each classification method are summarized in Table 5.10

Table 5. 10. Statistical versus CNN Methods on age prediction from Face Images and the time requirements to process 30 images

	Statistical Methods Best	**CNN Method Best**
Accuracy (%)	40.10 %	59.13%
Time (s)	271.71 s	16.63 s

As seen in Table 5.10, the CNN method has more accurate results than the statistical methods under the same conditions. The same dataset is used by applying the same pre-processing step (only image alignment to equalize the image size) on the images with the same evaluation technique. Consequently, it can be claimed that CNN is more powerful than the statistical methods in these conditions in terms of time periods to complete the process.

5.3 Age Prediction from 3D Body Data

Since 2002, age has been widely predicted using face images [151]. Nevertheless, there are some cases when face images cannot be obtained, and only the body portion can be captured. Under these conditions, our study proposes an artificial neural network (ANN) approach in order to predict the age range from body data. In order to train and test the system, we collected this data from 65 females and 105 males between 15 to 72 years old as volunteers in our experiment. In the collected data set, there are 3D coordinates of joint points from 103,982 scenes (or shots) belonging to 170 people and recorded for each scene. The procedure followed and their conditions have already been explained in Section 4.3.1, "Obtained Data from 3D Camera", and Section 4.3.2, "Data Expansion". Thus, the same body joint point coordinates are used in the gender and age prediction part of the study.

For the classification method, a multi-layer neural network is generated with the same input layers as 60 inputs (20 joint points with 3 coordinates, 20*3), the same sigmoid function in the hidden layers, softmax transfer function in the output layer with the

network that is trained for gender prediction, and the Levenberg-Marquardt training algorithm used to optimize weight, bias and error. However, the number of output layers is increased from two to six classes for the age ranges, as given in Table 5.11.

Table 5. 11. Age ranges, the classes of ANN and the number of people in the age range for age prediction

Age Range	Class	Number of People
< 20	1	12
20 – 30	2	47
30 – 40	3	33
40 – 50	4	28
50 – 60	5	34
60 <	6	16

According to the ANN created based on these information details, the number of hidden layers is changed from 10 to 60 (increased by 5 in each newly trained neural network).

For age prediction using body data, the trained ANNs are tested and validated both with 15% of all the scene data using whole body, upper body, and lower body data. The following subsections contain the experimental results of all the trained ANNs.

5.3.1 Classification Using All-Body Joint Points

As mentioned earlier, all the 103892 scene data are divided into three parts: 15% used for validation, 15% for testing and the remaining, and 70% to account for the training data. In the hidden layer, the number of neurons is changed starting from 10 to 60, increasing by fives, and the results are presented below. For *"Train, Test, Validation"* and *"All"*, the summation of all these three groups' accuracy percentages appear in Table 5.12.

According to the table, the highest accuracy is obtained from 60 neurons in the hidden layer for training, testing, validation and all of the dataset. By increasing the number of neurons in the hidden layer, the accuracy percentages also increase up to 99.0% in the training. Moreover, not only the training accuracy, but also the testing and validation accuracies support the training accuracy with slight differences among the percentages. This means that there is no overfitting. Additionally, the time requirement to process 30 images is calculated as 0.46 seconds for the 60 neurons in the hidden layer of the ANN. Also, 30 image joint points are tested using the pre-trained ANN with 60 neurons in the hidden layer.

Table 5. 12. ANN results of All-Body Joint Point Features

Number of Neurons in Hidden Layer	Train Accuracy %	Test Accuracy %	Validation Accuracy %	All Accuracy %
10	76.4	75.7	76.1	76.2
15	84.0	84.1	84.0	84.0
20	90.3	89.6	89.4	90.1
25	93.2	92.4	93.1	93.1
30	95.9	95.3	95.5	95.8
35	94.7	94.4	94.7	94.7
40	96.8	96.4	96.2	96.6
45	97.1	96.5	96.7	97.0
50	97.0	96.4	96.7	96.9
55	98.7	98.1	98.2	98.6
60	**99.0**	**98.5**	**98.3**	**98.8**

In this part, all the joint point features are used to determine the accuracy without decomposing any person's joint points. Using all the joint points for training may lead to positive bias in the experiments, to overcome which the *Leave-One-Person-Out* (LOPO) and *10-fold* cross validation methods are also applied.

Using LOPO, we test 170 people individually. For this test, one person is selected as the test case, and the body joint points of the remaining 169 people are used as the training data. This process is repeated for each person as being the test case. The confusion matrix of LOPO is presented in Table 5.13.

Table 5. 13. Confusion Matrix of LOPO for whole-body

		Actual					
		1	**2**	**3**	**4**	**5**	**6**
Predict	**1**	**12**	0	0	0	0	0
	2	0	**47**	0	0	**1**	0
	3	0	0	**33**	0	0	0
	4	0	0	0	**28**	0	0
	5	0	0	0	0	**33**	0
	6	0	0	0	0	0	**16**

As it is obviously seen in Table 5.13, among the 170 people, there is only one person, which is misclassified as in the age range between 20 and 30, whose correct age range is between 50 and 60. So the accuracy of LOPO is 99.41% which is good enough to predict age ranges from body joints points.

The 10-fold evaluation method is also applied for classification where 17 people are randomly selected for testing, and the remaining 153 are used as training data. The obtained results are given in Table 5.14.

Table 5. 14. Confusion Matrix of the 10-fold Evaluation Method

		Actual					
		1	**2**	**3**	**4**	**5**	**6**
Predict	**1**	**1**	0	0	0	0	0
	2	0	**4**	0	0	0	0
	3	0	1	**3**	0	0	0
	4	0	0	0	**3**	0	0
	5	0	0	**1**	0	**2**	**1**
	6	0	0	0	0	0	**1**

As shown the table, there are 17 tested samples from the dataset and 3 of them are misclassified. While the precise accuracy rate is 82.35%, the one-off accuracy rate is 94.11% calculated by taking into account the accuracy of the persons who are classified in the previous or next class with the actual class (1-class error case). Among 3 misclassified persons, 2 are predicted as one older range and one younger range, which can be accepted for certain age prediction applications. In all, the accuracy rate of the 10-fold classification method appears lower than the LOPO approach which can be explained as follows: in 10-fold, the size of the training set is decreased. It is known that in neural network applications if the size of the training data set is reduced, the training performance may deteriorate as well. A brief comparison between the LOPO and 10-fold cross validation results appears in Table 5.15.

Table 5. 15. Accuracies of LOPO and 10-fold technique for All-Body Joint Points

LOPO		**10-fold**	
Exact	**1-off**	**Exact**	**1-off**
99.41 %	99.41 %	94.11 %	82.35 %

As seen from Table 5.15, using the body information, the age ranges of the people can be predicted with high accuracy. However, in some cases, it is not possible to obtain

the full body data, to overcome which problem, upper- and lower-body joint points are used to predict the age range. The results are presented in the following sections.

5.3.2 Classification Using Upper-Body Joint Points

After all the body joint points are used in the previous part with promising accuracies, it was concluded that the upper body parts' joint points should also be considered to predict age ranges. In line with this purpose, only the first 11 of the joint points as stated in Table 4.7 are used as the upper body parts. With the same test, validation and training proportions of all scenes, this ANN process is repeated for the upper and lower body parts. The results of the accuracy rates for different neuron sizes in the hidden layer of ANN for upper body parts are shown in Table 5.16.

Table 5. 16. ANN results for Upper-Body Joint Points' Features

Number of Neurons in Hidden Layer	Train Accuracy %	Test Accuracy %	Validation Accuracy %	All Accuracy %
10	68.7	68.8	68.8	68.7
15	72.4	72.1	72.5	72.4
20	79.2	79.2	78.4	79.1
25	82.3	81.7	81.7	82.1
30	85.8	85.4	85.3	85.6
35	86.3	85.6	86.2	86.2
40	87.9	87.9	87.7	87.8
45	90.8	90.1	90.4	90.6
50	89.3	88.9	88.8	89.2
55	91.6	91.7	91.0	91.6
60	**92.5**	**92.1**	**91.8**	**92.3**

As seen in Table 5.16, 60 neurons in the hidden layer provide the highest accuracy rate for age prediction using the upper body parts of the data during the training, testing and validation process. Up to the 60 neurons in the hidden layer, increasing the neuron numbers also brought about an increase in the accuracy rate. This table shows the accuracy for each scene data with 70% of them used for training and 30% for validation

and testing. Each person in the data is tested using the LOPO method, and the confusion matrix for 170 people are tabulated as follows.

Table 5. 17. Confusion Matrix of LOPO for Upper Body Data

		Actual					
		1	**2**	**3**	**4**	**5**	**6**
Predict	**1**	**11**	0	0	0	0	0
	2	**1**	**47**	**1**	**2**	0	0
	3	0	0	**32**	0	**1**	0
	4	0	0	0	**26**	**1**	**1**
	5	0	0	0	0	**32**	0
	6	0	0	0	0	0	**15**

As seen in Table 5.17, among the 170 people, there are only seven people as misclassified. The actual accuracy of LOPO is 95.88 % for upper body joint points. Two in seven misclassified people are classified as in one younger age range; whereas one in seven misclassified people is classified as in one older age range. Therefore, the one-off accuracy rate is 97.65%, which is better than the actual accuracy rate.

5.3.3 Classification Using Lower Body Joint Points

The last 10 of the joint points, as indicated in Table 4.7, are used for the lower body part data as inputs for ANN, and the following Table, 5.18, states the accuracy rates of each ANN with different number of neurons in the hidden layer.

Table 5. 18 ANN results of the Lower-Body Joint Point Features

Number of Neurons in Hidden Layer	Train Accuracy %	Test Accuracy %	Validation Accuracy %	All Accuracy %
10	71.3	71.4	71.1	71.3
15	75.3	74.6	75.4	75.2
20	79.9	80.0	80.1	79.9
25	87.0	86.9	86.8	86.9
30	88.6	88.2	88.4	88.5
35	91.0	90.5	90.4	90.8
40	91.5	91.3	91.2	91.4
45	93.2	92.9	93.1	93.1
50	95.8	95.2	95.7	95.7
55	94.4	93.9	94.1	94.3
60	**96.1**	**95.7**	**95.6**	**96.0**

Accordingly, the best accuracy from the lower body joint points is obtained using 60 neurons in the hidden layer of ANN, similar with the previous findings presented in Tables 4.15 and 5.14. Once more, LOPO is used for classification and the confusion matrix is given in Table 5.19.

Table 5. 19. Confusion Matrix of LOPO for Lower Body Data

		Actual					
		1	**2**	**3**	**4**	**5**	**6**
Predict	**1**	**12**	0	0	0	0	0
	2	0	**47**	0	0	0	0
	3	0	0	**33**	0	**1**	0
	4	0	0	0	**27**	**2**	**1**
	5	0	0	0	**1**	**31**	0
	6	0	0	0	0	0	**15**

Based on this table, among 170 people, there are only five as incorrectly classified. The actual accuracy of LOPO is 97.05 % for the lower body joint points. Two in five misclassified people are classified as in one younger age range and one in five misclassified people is classified as in one older age range. Therefore, the one-off accuracy rate is 98.82%, which is better than the actual accuracy rate.

5.3.4 Age Prediction from 3D Body Data Using ANN-Conclusion

In this part of the study, a new perspective is offered to predict age based on 3D body data. The 3D coordinates of joint points are used as input to ANN, which is selected as a Multilayer Neural Network, and the number of neurons in the hidden layer is examined in the experiment to test and increase the accuracy values. In order to evaluate and verify the classifier accuracy, the input data is divided into three groups: training data, validation data and testing data. The data used in the classifier is collected from 170 volunteers ranging from 15 to 72 (Table 5.10). During the data collection, a depth camera is used and volunteers are free to take any position in front of the camera at will. Using the 3D skeleton data, the accuracy rates paint a promising picture in order to predict the age of people. The following table concludes the previous findings with the best accuracies so that it can be easily seen in the table - that using the information pertaining to the lower body parts of the body provides higher accuracies than the upper body parts' joint points. On the other hand, whole-body joint points give the highest accuracy among the others. In addition to this, the

LOPO and 10-fold evaluation methods also support the usability of 3D body coordinates of joint points to predict age.

Table 5. 20. Best accuracies for Upper, Lower and whole body joint points for all scene data

	Upper Body Joints	Lower Body Joints	Whole Body Joints
Train Accuracy %	92.5	96.1	99.0
Test Accuracy %	92.1	95.7	98.5
Validation Accuracy %	92.8	95.6	98.3

Table 5. 21. LOPO Results for Lower Body Joints, Upper Body Joints and Whole-Body Joints

	Lower Body	Upper Body	Whole Body
Exact Accuracy %	97.05	95.88	99.41
1-off Accuracy %	98.82	97.65	99.41

In this posture-based system, age can be predicted automatically using the 3D skeleton data with the help of a 3D camera. The main contribution of this study is that even if the camera cannot capture the face images or whole-body parts, the system obtains an accurate result for age prediction by using the upper or lower part of the human body. In crowed areas, capturing high-resolution face images and whole joint points of the body is difficult. For this reason, capturing only the joint points of the body parts should be enough for age prediction with our approach.

Our study of age prediction from 3D joint points is a novel approach because, so far, there have been no studies predicting people's age from 3D body coordinates. There are only two studies predicting age from body features, both proposed by Sandygulova et al.[72] and [152]. In both cases, effort is made to predict the age of children. The first one [72], uses the length of the head. The database of the study has 428 volunteer children between 5 and 18, and the mean absolute error is 0.94 with a standard deviation of 1.27. The second one [152], produced by the same authors, uses 3D skeleton data based on a set of motion sequences, and their database has 28 children between 7 and 16 with an accuracy rate 95.2%. Additionally, our method is proven to be more successful using 3D data captured from a significant number of real people.

CHAPTER 6

A SOFTWARE FOR GENDER AND AGE PREDICTION

Gender and age prediction applications are used in different areas, such as mobile applications [153], social media [154], text[74], browsing behavior[155] and face images [156]. Estimating age and gender from face images has been a research topic for many years, as already stated in Section 2, and the success in the field has been increasing day by day. Especially after the LAP Challenge competition [147], the populations of gender and age classification in appearance topics have been on the rise.

However, a software tool that estimates age and gender is yet to be developed in this field. So far, there have been some programs that estimates age and gender from images, such as *How old do I look?* [157], *Online Age Detector* [158] and *Kairos*[159], which estimate the age and gender of a person from the photo uploaded on the web. These three Internet-based age prediction products are stated as examples of automatic age detector [160]. The most commonly known web service *How old do I look?* Has been mentioned by some studies, one of which gives it as an example of age and gender prediction application using machine learning methods on the web [161]. Another one shows the image orientation sensitivity of it [162], and another mentions its large and balanced databases [163].

Although tests have been carried out related to the success of these programs, there has been no study of the availability of the programs in terms of age and gender prediction for automatic detection of a person on a video. It also appears that the existing products only predict age and gender from face images. Therefore, in this study, a software tool is developed to predict the age and gender of people from their face images or their body information. This software tool can be used to collect statistical information about the number of gender and age of the people is important. With the help of a developed software for this purpose, resources can be managed to predict the age group and their gender in shopping malls, stadiums, etc. It may be used

for human computer interaction for age- and gender-specific responses and contents. Additionally, it can be used for suspicious follow-up applications in security systems.

In this thesis, a software tool is presented for age and gender prediction. The details of the proposed tool is presented in Section 6.1 and the system usability scale tests of the proposed tool are presented in Section 6.2.

6.1 Graphical User Interface of the Proposed System

In our proposed system, gender and age can be predicted using the body data. The proposed system needs a face image for age, while gender estimates can be made using either face or whole-body images. Since the developed system is designed to obtain statistical data on age and gender using both 3D data and face images, no web application is developed. Also, there was no need to be in real-time because the system is produced for the purpose of providing statistical data. The following examples are given to illustrate each operation beginning with the first screen view of the developed system.

When the program starts, the following scene is seen in the window and the user can select the folder from the “Browse” button, where all the people’s images captured by the 3D camera are saved; otherwise, the pre-defined folder is assigned from the program.

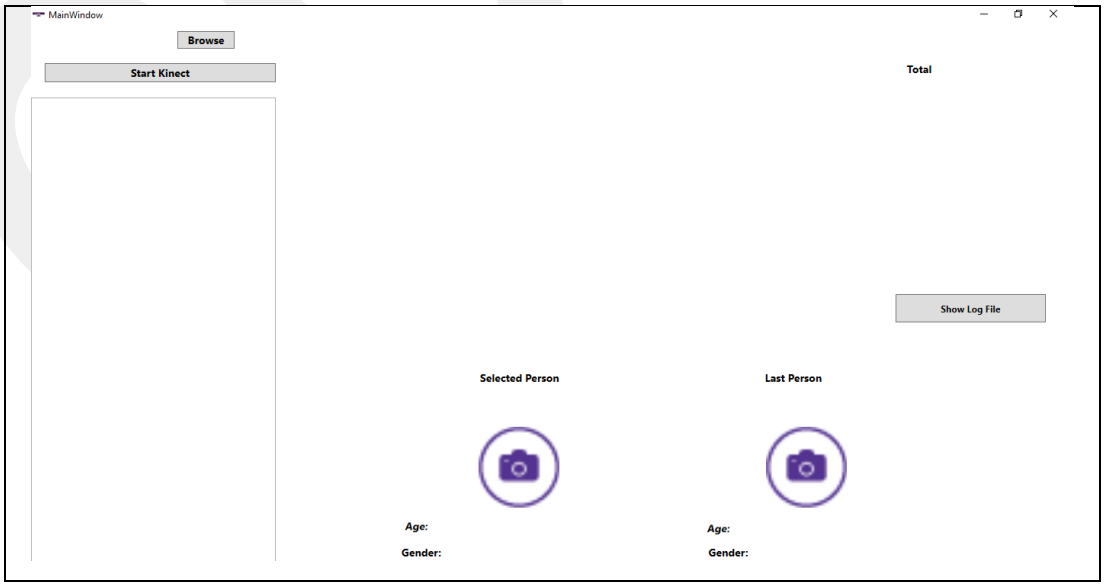


Figure 6. 1. The first opened state of the program

After folder selection, the "**Start Kinect**" button begins the capturing process and the camera scene image appears in the middle of the window. When a person comes in front of the camera, the joint points and the skeleton lines of the person is seen on the person image of the camera screen, as seen in the Figure 6.2.

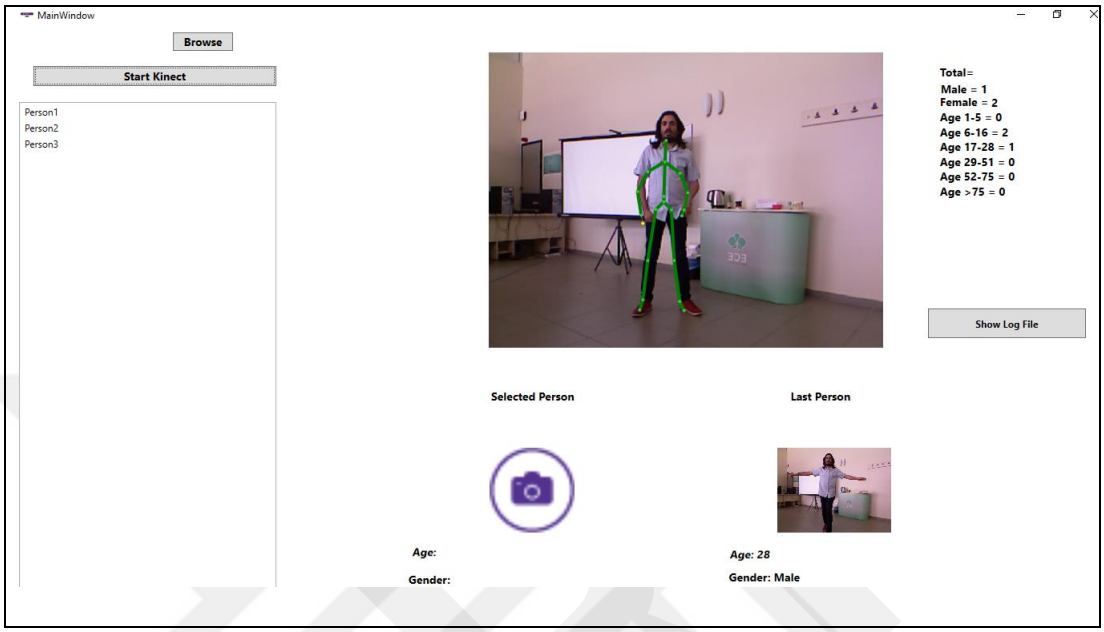


Figure 6. 2. A person and skeleton lines shown on the camera scene

The camera captures 30 frames in a second and fills the XML (Extensible Markup Language) file when the person is detected by the camera, until the person leaves. The structure of XML can be seen in Figure 6.3.

```
1    <?xml version="1.0" encoding="utf-8"?>
2    <ArrayOfCombineXMLInfo xmlns:xsi="http://www.w3.org/2001/X
3      <CombineXMLInfo>
4        <faceImages>
700      <heightInfo>
422      <jointsInfo>
484    </CombineXMLInfo>
485  </ArrayOfCombineXMLInfo>
```

Figure 6. 3. XML structure of the system

As seen in Figure 6.3, there are three parts in the XML structure for each person. In Figure 6.4, the "*faceImages*" header contains the frame name that the face is in, and four coordinates (right, left, top and bottom) of the rectangle of the face are all shown in an XML file. Those coordinate points provide the system to crop the face area in the captured image.

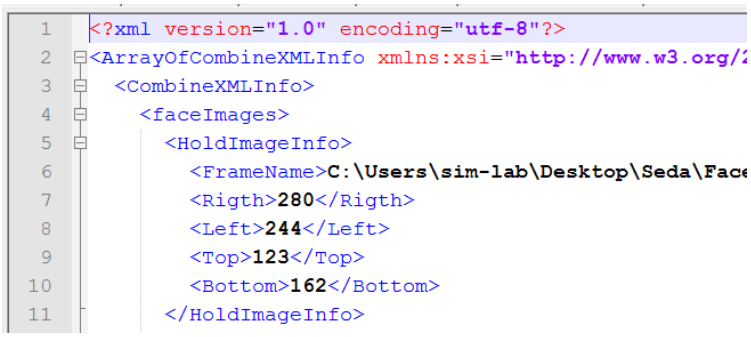


Figure 6. 4 Face area coordinate points' parts of the XML file

Another header part of the XML file is "*heightInfo*", which contains heights in meter in Figure 6.5. Three different calculated heights - skeleton, wingspan, and by estimation methods - are already explained in Section 4.2.3.

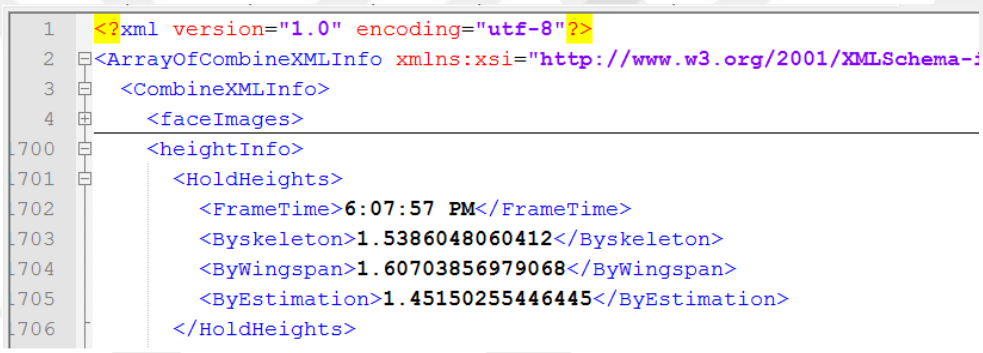


Figure 6. 5 XML file and calculated heights of a person

The final header in XML structure is "*jointsInfo*", which contains the frame time and 3D coordinates of each joint point. There are 20 joint points within the human body that are captured by Kinect v1, and each has x, y, and z coordinates stored as seen in Figure 6.6. In the figure, only the head, neck, left hand, left wrist, left elbow and left shoulder joints are shown with 3D coordinates.

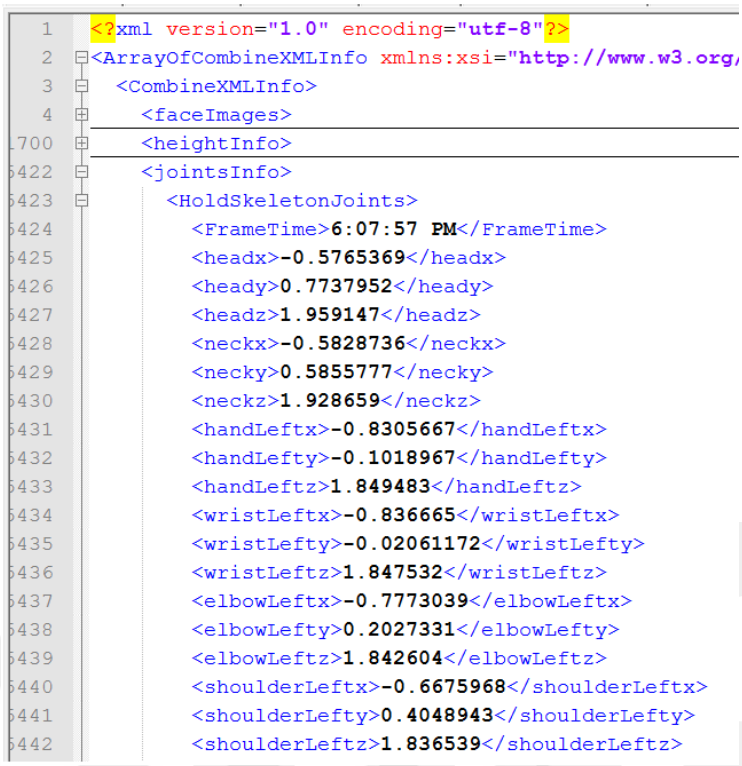


Figure 6. 6 Coordinates of some joint points

Along with the person leaving the screen, the completed XML file that is parsed to get the joint points coordinates crops the face area from the whole image and starts age and gender classification. This takes time, depending on the person standing in front of the camera. The number of saved joint point coordinates increase depending on the time duration that the person stands in front of the camera. Additionally, the number of face images captured by the system depends on the time duration that the person shows his face to the camera. For each skeleton, all joint point coordinates and each face image is classified individually as saved data in XML. The most repeated result for one person is sent to the user interface to show the predicted age and gender of the person.

On the interface, one of the images belongs to the last person shown at the bottom right of the screen with the age and gender information predicted by the system after the classification is completed. Moreover, the last person added to the end of the list is stated at the left side of the window. If one of the persons is selected in the list, an image and age-gender information is shown in the “Selected Person” label. All the statistical information about the people, who have been spotted by the camera up to that time, is represented with the gender and age range headers at the top-right side of the window. All these parts are shown in Figure 6.7 with their identification marks.

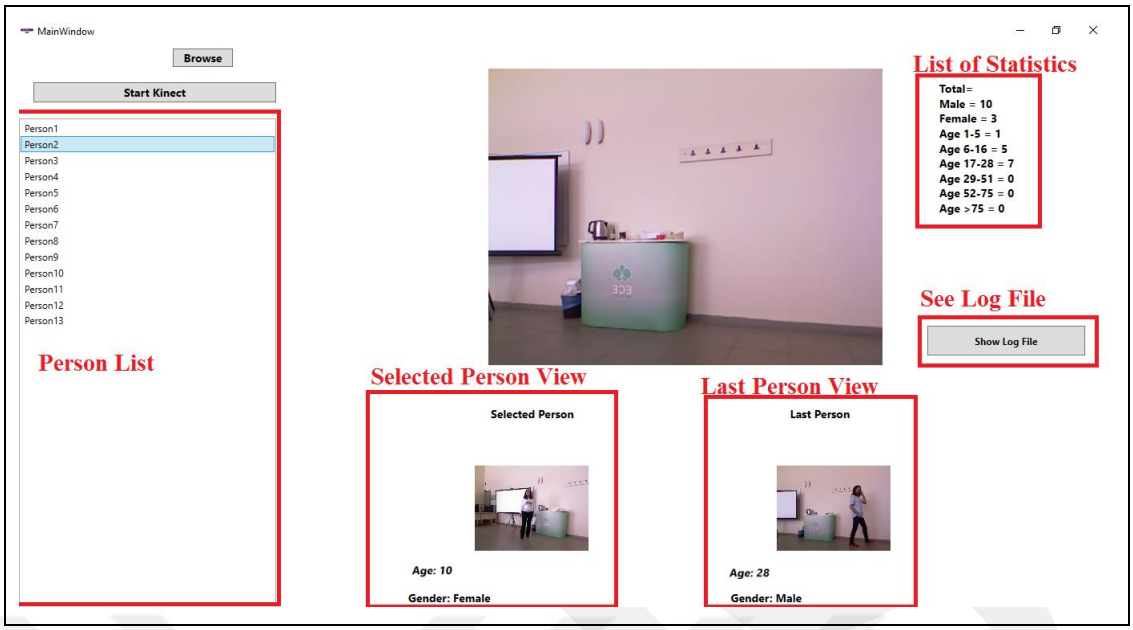


Figure 6. 7 Window with Identification Marks

After the gender and age prediction process is completed according to the data received from a person, all data related to age, gender, person serial numbers with the date and time of that moment are recorded in the log in the form of a text file. To see all the logs, the "Show Log file" button can be pressed and the text file is opened, as in Figure 6.8.

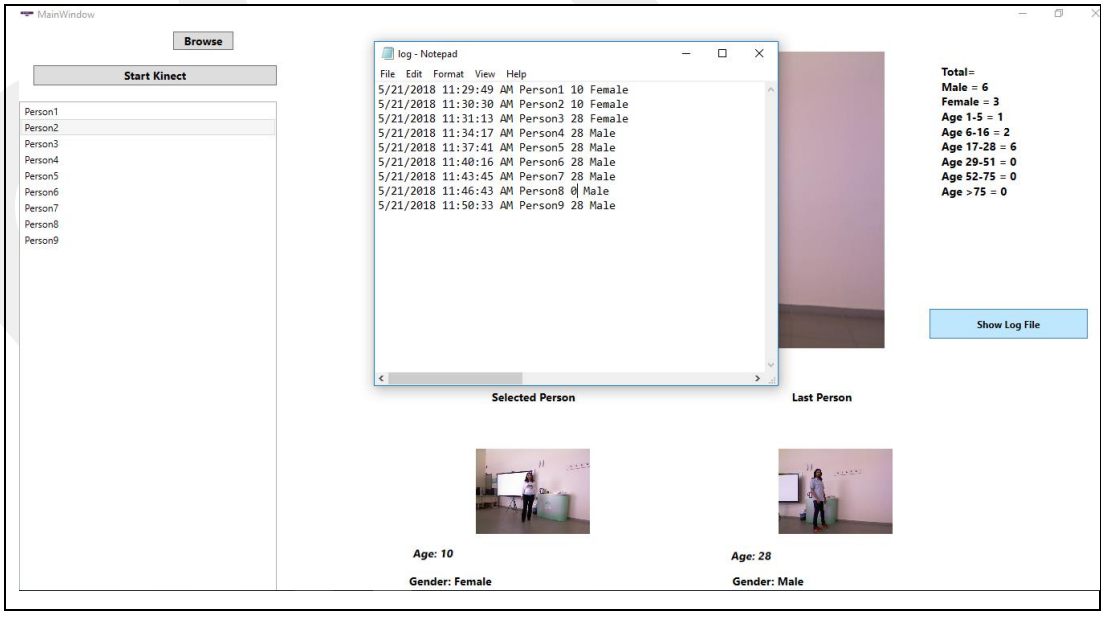


Figure 6. 8 The Log file text of the system

This system provides age and gender prediction according to the captured human data from a camera which provides 3D body information and RGB images. The software

is pure, simple to use and designed to be goal-oriented as explained with images shown earlier.

6.2 System Usability Scale Test Results

To prove the usability of the system software, it is tested by 86 people, all of whom were asked to respond to a SUS (System Usability Scale) questionnaire as per Table 6.1. There are 25 females and 61 males between 21 and 41. The distribution of age and gender of the participants are presented in Figure 6.9.

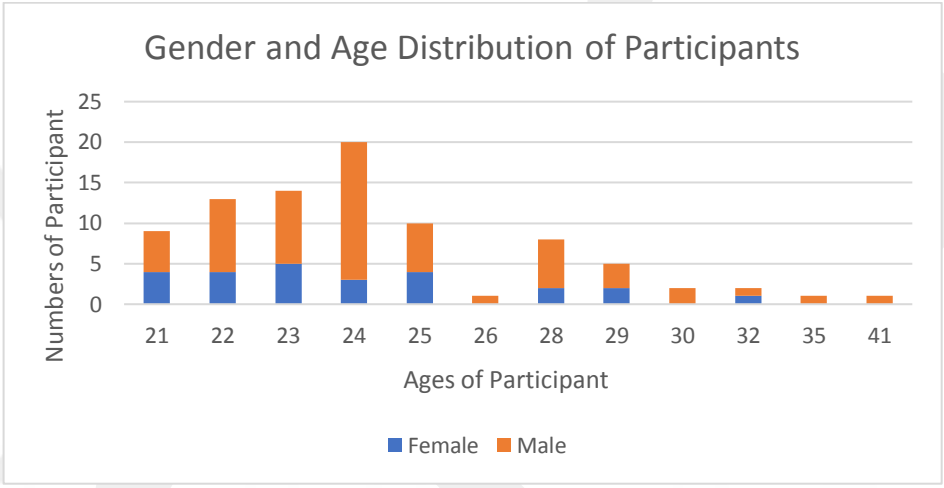


Figure 6. 9. Gender and age distribution of the SUS participants

As seen in Figure 6.9, most of the participants are male, between 21 and 25, and students in the field of Computer Engineering (32 people), Software Engineering (36 people) and Information Systems Engineering (17 people). The reason they were asked to do the tests is that they should be able to comment on the comfortable and convenient use and design of this software.

Table 6.2 shows the items of the scale and the number of the selected Likert items for each scale which change between "*Strongly Disagree*" and "*Strongly Agree*" for each scale item, as below.

The first scale is about the frequency of use of the system, and 43% of the participants agree, 28% of them strongly agree and only 11% of them disagree or strongly disagree with this scale. This means that the system is suitable to be used frequently according to the respondents.

The second scale tests the complexity of the system; 42% of the participants did not find the system to be complex and disagreed with this scale meaning; whereas, 24%

of them strongly disagree. Only 14% of the participants agree or strongly agree with the idea that this system is unnecessarily complex.

Table 6. 1 System Usability Scale

	Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
1. I think that I would like to use this system frequently	5	4	16	**37**	24
2. I found the system unnecessarily complex	21	**36**	17	8	4
3. I thought the system was easy to use	3	1	5	35	**42**
4. I think that I would need the support of a technical person to be able to use this system	14	21	18	**27**	6
5. I found the various functions in this system were well integrated	0	2	20	**43**	21
6. I thought there was too much inconsistency in this system	14	25	**27**	12	8
7. I would imagine that most people would learn to use this system very quickly	2	4	6	36	**38**
8. I found the system very cumbersome to use	14	19	**33**	11	9
9. I felt very confident using the system	2	3	15	**36**	30
10. I needed to learn a lot of things before I could get going with this system	19	**23**	17	19	8

In terms of the ease of using the system, third criterion shows that 49% of the participants strongly agree, and 41% of them agree. A very small group of 4% indicated that they were not in agreement. This scale confirms that our system is very easy to use.

In terms of the system's ease-of-use, the third criterion shows that 49% of the participants strongly agree, and 41% agree. A very small group (4%) indicated that they were not in agreement. This scale confirms that our system is very easy to use.
In the participants' set, the observed distribution in the fourth scale does not concentrate on "agree or disagree". However, the 'agree', 'neutral' and 'disagree' option percentages are 31%, 21%, and 25%, respectively, which are very close to each other compared with other scales. As a result, some people do not need technical staff to use the system, but some of them required it.

In the fifth scale, the system functions' integration is inquired, and 50% of the participants agree with the well-integration of the functions and 25% strongly agree. There is no strong disagreement and only 2% disagreement. This scale shows that the success of integrating the functions of the system is high.

32% of the participants thought that the system is neither consistent nor inconsistent (neutral), but 29% of the other participants disagreed that is, they found the system inconsistent. Moreover, the percentage of agreement and strong agreement of the inconsistency of the system is totally 23%, which is not negligible. Therefore, the participants' agreement or disagreement percentage is not obtained, which makes a difference in inconsistency. The reason for this incoherence on inconsistency derives from the fact that age estimation is not very successful, from the answers given to the open-ended questions in the proposals.

44% of the participants strongly agree and 42% agree that they learned how to use this system very quickly. Hence, the seventh criterion provides data as evidence that the system can be learned quite easily by users.

In the eight scale, the participants test if the system is cumbersome to use; the 'neutral' item is selected with 38% percentage value, while other items selection percentages are very close to each other, such as 22% disagree, 16% strongly disagree, 13% agree, and 11% strongly agree. As it is understood from the open-ended question, the participants thought that the process time of the system made it cumbersome.

42% of the participants state that they feel confident while using the system and 35% of all strongly agree; only 6% disagree or strongly disagree. This result shows that people do not feel threatened while using the system.

The last scale tests the necessity of training people before using the system. 27% is the highest value of percentage disagreeing with the idea, 9% is the lowest percentage with strong agreement and the rest has nearly equal percentages.

The pie chart representation of each scale is presented in the appendix - SUS Pie Chart Representation Section- to reach the percentage of each Likert item for the scales.

According to the SUS, the strengths of the system in terms of the results we receive from the participants can be summarized as follows:

- ✓ The system can be used frequently (71% agree and strongly agree)
- ✓ The system is unnecessarily complex (66% disagree and strongly disagree)
- ✓ The system is easy to use (90% agree and strongly agree)
- ✓ The functions in the system are well-integrated (75% agree and strongly agree)
- ✓ It is easy to learn to use this system very quickly (84% agree and strongly agree)
- ✓ People feel confident while using the system (77% agree and strongly agree)
- ✓ It requires training those who use the system (49% disagree and strongly disagree)

As for the aspects to be strengthened, they include:

- ❖ Technical support to use the system (38% agree and strongly agree)
- ❖ Inconsistency in the system (23% agree and strongly agree)
- ❖ Very cumbersome system (24% agree and strongly agree)

To improve the system according to the responses, the followings are offered:

- ➢ A manual to describe the system and how to use the system can be added to the interface,
- ➢ The age classification part should be improved and re-integrated into the system,
- ➢ The response time should be shortened.

Overall, it can be said that the availability of the system is high, but we are aware that there are still other aspects to be improved and, hence, considered as future plans.

CHAPTER 7

DISCUSSION AND CONCLUSION

In this study, automatic gender and age prediction methods that use face images and 3D body information are examined and a software tool is generated to predict age and gender of people. For this purpose, first gender prediction from face images is considered on the FERET database using LBP and HOG statistical methods to extract face image features alongside SVM and KNN classification methods. The experimental results show that the SVM classifier is more effective than KNN, and that the combination of LBP and HOG features with the SVM classifier gives the best accuracy for gender prediction. Previous studies proposed by Khalifa [14], Nazir [59], Shih [61], Liu [62] all used other statistical methods, such as DCT (Discreate Cosine Transform), and Precise Patch Histogram (PPH), obtaining high accuracy values such as 94%. In this study, gender prediction from face images gives reasonable results. In general, if face images are captured with a 3D camera, gender can be predicted easily using these images. Otherwise, that is if face images cannot be captured or the resolution of the captured image is low, then gender prediction cannot be done. The solution for this is to extract body features, such as anthropometric measurements, to predict gender.

Gender prediction is also examined with body information using two classification methods: statistical methods and the neural network approach. The body information is used to extract anthropometrics measurements and statistical SVM and KNN methods and ANN methods are applied on these extracted features. The best accuracy is 96.77% obtained using SVM with the MLP kernel function. Moreover, upper body, lower body and whole-body information are separately examined and gender prediction accuracy from such body information is improved up to 99.26 %, 97.83 % and 97.29%, respectively with the LOPO (or Leave One Person Out) evaluation method for the ANN approach using the joint points' coordinates. Kakadiaris et al.

[15] carried out a study which achieved 95% accuracy using the gait and anthropometric measurements. The present work manages to achieve 96.77% accuracy using only the anthropometric measurements. Additionally, Kakadiaris proposed a study using upper and lower body parts using a Kinect camera, and the accuracy values were 86.0% all body, 78.0% lower body, and 72.0% upper body for real images. The Kakadiaris method is more powerful in the CAESAR database, whose data is different from that captured by Kinect.

When we compare our study with these previous attempts about using upper body information to predict gender (accuracies at 86.0% all body, 78.0% lower body, and 72.0% upper body for real images) [15], it can be seen that the accuracy achieved in the present work is higher. In addition, this higher accuracy is achieved with a set of data generated from more people when compared to previous works in the literature. As a result, our contribution to gender prediction from body information is the improvement of the accuracy and using the upper and lower body parts rather than the whole body. Only the upper part or the lower part of the body features also give better accuracy rates in our study than the previous ones [15]. Therefore, it is established here that only one part of the body features is enough to predict gender, and that the whole-body information is not mandatory in our proposed method. Comparing to previous studies [15], our method has also been tested with more measurements, thereby revealing that the proposed method is more general than the previous ones. Lastly, in our study, gender prediction using body parts of the people is done in a short time, such as 0.12 seconds, which is nearly real-time with high accuracies.

Another main objective of the study was age prediction using face images and body information. To achieve this goal, statistical pattern recognition and convolutional neural network techniques are applied on the Gallagher's database, which is used for training face images because of dimension similarities captured with an RGB camera. For example, images in the FG-Net database are 640 x 480 with high resolution in general, but it is not always possible to capture such quality snapshots for faces. In situations where small-size images are not needed, the FG-NET data set can give an accuracy rate of around 99.86% [14] using statistical methods. According to the experimental results, the best accuracy rate is obtained at 40.1% for statistical methods. There are more accurate results for age prediction on Gallagher's as the site some additional feature extraction and classification methods with different evaluation methods as is mentioned in Section 5.2. In addition to statistical methods,

convolutional neural network (CNN) is used to predict age on the same dataset. Here, age prediction accuracy rate from face images is increased to 59.1% using the CNN technique. These two statistical and CNN models are both tested on 30% of the Gallagher's dataset, and 70% of them are used for training.

In this study, statistical methods (LBP, HOG, SVM, and KNN) are compared with popular deep learning techniques, such as CNN, in order to predict age from face images. The result of the CNN method is greater than that of the statistical method as mentioned above. The differences between these methods start with the procedures to be followed; both statistical and CNN methods have the pre-processing phase to resize the images in order to have an equal number of features. In the statistical methods, one of the feature extraction methods is needed to extract features; however, features are extracted in the convolutional layers in CNN, where there is more than one convolutional layer in deep learning, making feature extraction possible more than once and according to the weights of the neurons. Moreover, back propagation mechanism in CNN makes it possible to update the weights of the neurons according to the validation accuracy. However, extracting face features occurs only once before the classification in the statistical methods. There are both advantages and disadvantages in each method. To begin with, the training time of a CNN is longer than the statistical methods because the feature selection process occurs during the training time of CNN. However, there is no feature selection process in statistical methods if it is not added following to the feature extraction process. On the other hand, if the amount of the training data is less for CNN, overfitting occurs, which means that the system memorizes the training set and, hence, different data cannot be classified correctly. On the contrary, statistical methods do not large amounts of training data to train the system, which is an advantage. In addition to these, time requirement for prediction process of CNN (16.63 s) is less than the statistical methods (271.71 s) in the condition which provides the best accuracy (LBP+SVM).

Age prediction from body information is also examined on our own dataset with the ANN technique consisting of the joint points' coordinates related with posture to offer data about age [164]. The accuracy results based on LOPO for whole-body, upper body part and lower body part are 99.41%, 95.88% and 97.05% respectively.

The contribution of our study is using 3D body metrics to predict age. Previous body-based [165] and gait-based [166] age prediction methods depend on 2D images. Additionally, these studies need either a walking step period of people or whole-body

image to predict ages. There are only two studies that predict age and gender from 3D body information; one of them [72] considers the head height and the other one [152] the 3D motion; yet, both use children in their dataset. In the Sandygulova's first study [72], the length of the children's head is used to predict the age. The database of the study has 428 volunteer children between 5 and 18 and the mean absolute error is 0.94 with a standard deviation of 1.27. In the Sandygulova's second study [152] the 3D skeleton data based on set of motion sequences is used and their database has 28 children between 7 and 16 with an accuracy rate of 95.2%. However, those studies do not depend on adult body information in contrast to ours. There is no study which uses 3D body information for adults in the literature. In the present work, in addition, age prediction is done in 0.46 second with high accuracies from the joint point coordinates of body parts for adults, a first attempt in the literature. Another point is that, in our study, age can be predicted with a high accuracy rate without using face images and upper-body and lower-body parts are considered, instead, to predict age with high accuracy rates.

Lastly, the face and body methods are combined in a software tool to predict age and gender with integrated methods in the software tool and using a 3D camera. No previous study in the literature has ever attempted such an integrated system. To test and validate, the developed software is tested for usability by senior students of Computer Engineering, Software Engineering and Information Systems Engineering departments.

To sum up, in this study, a software tool to predict age and gender based on both face and body information is developed and studied with its components in detail. It is also determined from the study that CNN is more effective than statistical methods for age prediction from face images. When such images could not be captured with a camera, using the body information instead can realize age and gender prediction with high accuracy rates.

CHAPTER 8

LIMITATIONS AND FUTURE WORK

There are certain limitations in terms of data collection as an ethical committee report was published and permission to receive data from individuals was taken upon their consent. However, it is difficult to collect data from people who are less than 18 years old as both parents must approve. Here, there are no participants under 15 in the dataset. Additionally, there are no people of any other race or nationality than white Turkish because the data was collected at Atilim University, where multi-ethnicity or multi-nationality is a rather rare phenomenon. As such, one of the future plans of the study is to increase the number of individuals in the dataset and to extend the age range as much as possible.

Another limitation in collecting data is that each individual person faces the camera while the face and body information is captured. Thereby, in the proposed system and developed software, only one person is tested at a time. In further studies, it will be taken it to a public place to predict age and gender of people using face images, 3D body information, and anthropometric measurements.

A third limitation is considering the body information in three phases whole, upper and lower parts. Nevertheless, some other sub-parts of the body may be used to predict gender and age. For example, face joint points coordinates [167][168] from 3D camera, or head movements [169] can be used for this purpose; not to mention to estimate the age and gender by considering the right or left side of the body as future work and objectives.

In the future, comparison of deep learning and statistical methods can be done on different databases to predict age and gender. In this study, only LBP, HOG, SVM and KNN statistical methods have been used to predict age and gender; however, there are several statistical methods in the literature which can be tested for predicting age and gender. Additionally, the differences between transfer learning and standard learning in convolutional neural network (CNN) can be examined for age and gender

prediction. Furthermore, there is no distinction as to whether or not a person is wearing makeup. Still, datasets available on the web have been used in predict age and gender regardless of the presence of makeup on users' faces. Therefore, no study has been done to determine whether a person with makeup will change their age and gender estimates. To tell the difference and determine the related factors, a future work in this regard can be useful as well.

REFERENCES

[1] D. Steffensmeier, N. Painter-Davis, and J. Ulmer, "Intersectionality of race, ethnicity, gender, and age on criminal punishment," *Sociol. Perspect.*, vol. 60, no. 4, pp. 810–833, 2017.

[2] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "Dager: Deep age, gender and emotion recognition using convolutional neural network," *arXiv Prepr. arXiv1702.04280*, 2017.

[3] N. Bansal, A. Verma, I. Kaur, and D. Sharma, "Multimodal biometrics by fusion for security using genetic algorithm," in *Signal Processing, Computing and Control (ISPCC), 2017 4th International Conference on*, 2017, pp. 159–162.

[4] B.-K. Park, J. C. Lumeng, C. N. Lumeng, S. M. Ebert, and M. P. Reed, "Child body shape measurement using depth cameras and a statistical body shape model," *Ergonomics*, vol. 58, no. 2, pp. 301–309, 2015.

[5] M. Ålgars *et al.*, "The adult body: How age, gender, and body mass index are related to body image," *J. Aging Health*, vol. 21, no. 8, pp. 1112–1132, 2009.

[6] S. X. M. Yang, P. K. Larsen, T. Alkjær, B. Juul-Kristensen, E. B. Simonsen, and N. Lynnerup, "Height estimations based on eye measurements throughout a gait cycle," *Forensic Sci. Int.*, vol. 236, pp. 170–174, 2014.

[7] Q. Riaz, A. Vögele, B. Krüger, and A. Weber, "One small step for a man: estimation of gender, age and height from recordings of one step by a single inertial sensor," *Sensors*, vol. 15, no. 12, pp. 31999–32019, 2015.

[8] C. Pfitzner, S. May, and A. Nüchter, "Evaluation of Features from RGB-D Data for Human Body Weight Estimation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10148–10153, 2017.

[9] C. Pfitzner, S. May, and A. Nüchter, "Body Weight Estimation for Dose-Finding and Health Monitoring of Lying, Standing and Walking Patients Based on RGB-D Data," *Sensors (Basel).*, vol. 18, no. 5, 2018.

[10] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 3, pp. 441–467, 2016.

[11] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognit. Lett.*, vol. 68, pp. 218–230, 2015.

[12] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Computer Vision and Pattern Recognition, 2009. CVPR*

2009. IEEE Conference on, 2009, pp. 112–119.

[13] E. Gonzalez-Sosa, A. Dantcheva, R. Vera-Rodriguez, J.-L. Dugelay, F. Brémond, and J. Fierrez, "Image-based gender estimation from body and face across distances," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*, 2016, pp. 3061–3066.

[14] T. A. M. KHALIFA, "PREDICTING AGE AND GENDER OF PEOPLE BY USING IMAGE PROCESSING TECHNIQUES," 2016.

[15] I. A. Kakadiaris, N. Sarafianos, and C. Nikou, "Show me your body: Gender classification from still images," in *Image Processing (ICIP), 2016 IEEE International Conference on*, 2016, pp. 3156–3160.

[16] D. Cao, C. Chen, D. Adjeroh, and A. Ross, "Predicting gender and weight from human metrology using a copula model," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, 2012, pp. 162–169.

[17] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proceedings of the 16th ACM international conference on Multimedia*, 2008, pp. 725–728.

[18] G. Guo, G. Mu, and Y. Fu, "Gender from body: A biologically-inspired approach with manifold learning," in *Asian Conference on Computer Vision*, 2009, pp. 236–245.

[19] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 2009, pp. 1235–1242.

[20] A. S. Won, L. Yu, J. H. Janssen, and J. N. Bailenson, "Tracking gesture to detect gender," in *Proceedings of the International Society for Presence Research Annual Conference*, 2012, pp. 24–26.

[21] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1–8.

[22] V. E. Kelly, M. A. Schrager, R. Price, L. Ferrucci, and A. Shumway-Cook, "Age-associated effects of a concurrent cognitive task on gait speed and stability during narrow-base walking," *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.*, vol. 63, no. 12, pp. 1329–1334, 2008.

[23] P. C. Grabiner, S. T. Biswas, and M. D. Grabiner, "Age-related changes in spatial and temporal gait variables," *Arch. Phys. Med. Rehabil.*, vol. 82, no. 1, pp. 31–35, 2001.

[24] Q. Xiao, "A biometric authentication approach for high security ad-hoc networks," in *Information Assurance Workshop, 2004. Proceedings from the Fifth Annual IEEE SMC*, 2004, pp. 250–256.

[25] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Trans. Inf. forensics Secur.*, vol. 9, no. 3, pp. 464–475, 2014.

[26] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: a multiresolution convolutional

neural network for ECG-based biometric human identification in smart health applications," *IEEE Access*, vol. 5, pp. 11805–11816, 2017.

[27] A. Anand, R. D. Labati, M. Hanmandlu, V. Piuri, and F. Scotti, "Text-independent speaker recognition for Ambient Intelligence applications by using Information Set Features," in *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2017 IEEE International Conference on*, 2017, pp. 30–35.

[28] L. Geng, K. Zhang, X. Wei, and X. Feng, "Soft Biometrics in Online Social Networks: A Case Study on Twitter User Gender Recognition," in *Applications of Computer Vision Workshops (WACVW), 2017 IEEE Winter*, 2017, pp. 1–8.

[29] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2401–2412, 2013.

[30] Q. You, S. Bhatia, T. Sun, and J. Luo, "The eyes of the beholder: Gender prediction using images posted in online social networks," in *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, 2014, pp. 1026–1030.

[31] J. Zhang *et al.*, "Reliable Gender Prediction Based on Users' Video Viewing Behavior," in *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, 2016, pp. 649–658.

[32] D. Duong, H. Tan, and S. Pham, "Customer gender prediction based on E-commerce data," in *Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on*, 2016, pp. 91–95.

[33] M. Topaloglu and S. Ekmekci, "Gender detection and identifying one's handwriting with handwriting analysis," *Expert Syst. Appl.*, vol. 79, pp. 236–243, 2017.

[34] A. Dantcheva, P. Elia, and A. Ross, "What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics," *IEEE Trans. Inf. Forensics Secur.*, 2016.

[35] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Multimodal gender detection," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 302–311.

[36] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti, "Your installed apps reveal your gender and more!," in *Proceedings of the ACM MobiCom workshop on Security and privacy in mobile environments*, 2014, pp. 1–6.

[37] Q. Wu and G. Guo, "Gender recognition from unconstrained and articulated human body," *Sci. World J.*, vol. 2014, 2014.

[38] F. Lin, Y. Wu, Y. Zhuang, X. Long, and W. Xu, "Human gender classification: a review," *Int. J. Biom.*, vol. 8, no. 3–4, pp. 275–300, 2016.

[39] A. Jalal, S. Kamal, and D. Kim, "Shape and motion features approach for activity tracking and recognition from kinect video camera," in *Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on*, 2015, pp. 445–450.

[40] A. Farooq, A. Jalal, and S. Kamal, “Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map.,” *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 5, 2015.

[41] H. Han, C. Otto, X. Liu, and A. K. Jain, “Demographic estimation from face images: Human vs. machine performance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1148–1161, 2015.

[42] Y. Andreu, P. García-Sevilla, and R. A. Mollineda, “Face gender classification: A statistical study when neutral and distorted faces are combined for training and testing purposes,” *Image Vis. Comput.*, vol. 32, no. 1, pp. 27–36, 2014.

[43] G. L. Farina, F. Spataro, A. De Lorenzo, and H. Lukaski, “A Smartphone Application for Personal Assessments of Body Composition and Phenotyping,” *Sensors*, vol. 16, no. 12, p. 2163, 2016.

[44] A. Scano, A. Chiavenna, M. Malosio, and L. Molinari Tosatti, “Kinect V2 Performance Assessment in Daily-Life Gestures: Cohort Study on Healthy Subjects for a Reference Database for Automated Instrumental Evaluations on Neurological Patients,” *Appl. Bionics Biomech.*, vol. 2017, 2017.

[45] R. Buffa *et al.*, “A new, effective and low-cost three-dimensional approach for the estimation of upper-limb volume,” *sensors*, vol. 15, no. 6, pp. 12342–12357, 2015.

[46] A. Skalski and B. Machura, “Metrological analysis of microsoft kinect in the context of object localization,” *Metrol. Meas. Syst.*, vol. 22, no. 4, pp. 469–478, 2015.

[47] E. Makinen and R. Raisamo, “Evaluation of gender classification methods with automatically detected and aligned faces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, 2008.

[48] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, “Sexnet: A neural network identifies sex from human faces.,” in *NIPS*, 1990, vol. 1, p. 2.

[49] S. S. Liew, M. K. Hani, S. A. Radzi, and R. Bakhteri, “Gender classification: a convolutional neural network approach,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, no. 3, pp. 1248–1264, 2016.

[50] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von der Malsburg, “Face recognition and gender determination,” 1995.

[51] R. Brunelli and T. Poggio, “Face recognition: Features versus templates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.

[52] S. Gutta, H. Wechsler, and P. J. Phillips, “Gender and ethnic classification of face images,” in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 194–199.

[53] B. Moghaddam and M.-H. Yang, “Gender classification with support vector machines,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 306–311.

[54] A. Jain, J. Huang, and S. Fang, “Gender identification using frontal facial images,” in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, 2005, p. 4–pp.

[55] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[56] C.-C. Lai, C.-H. Wu, S.-T. Pan, S.-J. Lee, and B.-H. Lin, "Gender Recognition Using Local Block Difference Pattern," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, Springer, 2017, pp. 45–52.

[57] P. Rai and P. Khanna, "A gender classification system robust to occlusion using Gabor features based (2D) 2 PCA," *J. Vis. Commun. Image Represent.*, vol. 25, no. 5, pp. 1118–1129, 2014.

[58] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao, "Gender classification based on boosting local binary pattern," in *International Symposium on Neural Networks*, 2006, pp. 194–201.

[59] H.-C. Lian and B.-L. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *International Symposium on Neural Networks*, 2006, pp. 202–209.

[60] M. Nazir, M. Ishtiaq, A. Batool, M. A. Jaffar, and A. M. Mirza, "Feature selection for efficient gender classification," in *Proceedings of the 11th WSEAS International Conference*, 2010, pp. 70–75.

[61] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[62] H.-C. Shih, "Robust gender classification using a precise patch histogram," *Pattern Recognit.*, vol. 46, no. 2, pp. 519–528, 2013.

[63] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[64] H. Liu, Y. Gao, and C. Wang, "Gender identification in unconstrained scenarios using self-similarity of gradients features," in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 5911–5915.

[65] T. KHALIFA and G. ŞENGÜL, "YEREL İKİLİ ÖRÜNTÜ VE YÖNLÜ GRADYANT HİSTOGRAMI KULLANILARAK YÜZ GÖRÜNTÜLERİNDEN CİNSİYET TAHMİNİ," *Ömer Halisdemir Üniversitesi Mühendislik Bilim. Derg.*, vol. 7, no. 1, pp. 14–22, 2018.

[66] J.-H. Yoo, D. Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," in *ACIVS*, 2005, vol. 5, pp. 138–145.

[67] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 155–162.

[68] Y. Makihara, H. Mannami, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," in *Asian Conference on Computer Vision*, 2010, pp. 440–451.

[69] R. Miyamoto and R. Aoki, "Gender prediction by gait analysis based on time series variation on joint position," *J. Syst. Cybern. Informatics*, vol. 13, no. 3, pp. 75–82, 2015.

[70] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *Information forensics and security (WIFS), 2010 IEEE international workshop on*, 2010, pp. 1–6.

[71] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming, "Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report. Volume 1. Summary," SYTRONICS INC DAYTON OH, 2002.

[72] A. Sandygulova, M. Dragone, and G. M. P. O'Hare, "Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, 2014, pp. 826–831.

[73] V. O. Andersson, L. S. Amaral, A. R. Tonini, and R. M. Araujo, "Gender and Body Mass Index Classification Using a Microsoft Kinect Sensor.," in *FLAIRS Conference*, 2015, pp. 103–106.

[74] D. Nguyen, N. A. Smith, and C. P. Rosé, "Author age prediction from text using linear regression," in *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 2011, pp. 115–123.

[75] J. K. Y. Cheng, A. R. L. Fernandez, R. G. M. M. Quindoza, S. E. Tan, and C. Cheng, "Age and Gender Profiling of Social Media Accounts," *Age (Omaha).*, 2017.

[76] A. Smith and M. Gaur, "What's my age?: Predicting Twitter User's Age using Influential Friend Network and DBpedia," *arXiv Prepr. arXiv1804.03362*, 2018.

[77] S. Rosenthal and K. McKeown, "Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 763–772.

[78] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, no. 2–4, pp. 144–157, 2018.

[79] A. M. Albert, K. Ricanek Jr, and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Sci. Int.*, vol. 172, no. 1, pp. 1–9, 2007.

[80] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, 2008.

[81] A. Lanitis, "Facial age estimation," *Scholarpedia*, vol. 5, no. 1, p. 9701, 2010.

[82] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognit.*, vol. 46, no. 3, pp. 628–641, 2013.

[83] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the FG-NET ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, 2016.

[84] J. Liu, Y. Ma, L. Duan, F. Wang, and Y. Liu, “Hybrid constraint SVR for facial age estimation,” *Signal Processing*, vol. 94, pp. 576–582, 2014.

[85] D. S. Modha and Y. Fainman, “A learning law for density estimation,” *IEEE Trans. neural networks*, vol. 5, no. 3, pp. 519–523, 1994.

[86] K. Ricanek and T. Tesafaye, “Morph: A longitudinal image database of normal adult age-progression,” in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 341–345.

[87] A. C. Gallagher and T. Chen, “Understanding images of groups of people,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 256–263.

[88] E. Fazl-Ersi, M. E. Mousa-Pasandi, R. Laganiere, and M. Awad, “Age and gender recognition using informative features of various types,” in *Image processing (icip), 2014 ieee international conference on*, 2014, pp. 5891–5895.

[89] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of unfiltered faces,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 12, pp. 2170–2179, 2014.

[90] L. Wolf, T. Hassner, and Y. Taigman, “Descriptor based methods in the wild,” in *Workshop on faces in’real-life’images: Detection, alignment, and recognition*, 2008.

[91] C. M. Bishop, “Pattern recognition and machine learning (information science and statistics),” 2006.

[92] R. Azarmehr, R. Laganiere, W.-S. Lee, C. Xu, and D. Laroche, “Real-time embedded age and gender classification in unconstrained video.,” in *CVPR Workshops*, 2015, pp. 56–64.

[93] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[94] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, “The FERET database and evaluation procedure for face-recognition algorithms,” *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.

[95] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, “Robust face detection using the hausdorff distance,” in *International Conference on Audio-and Video-Based Biometric Person Authentication*, 2001, pp. 90–95.

[96] M. Minear and D. C. Park, “A lifespan database of adult facial stimuli,” *Behav. Res. Methods, Instruments, Comput.*, vol. 36, no. 4, pp. 630–633, 2004.

[97] J. Brooke, “SUS-A quick and dirty usability scale,” *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.

[98] M. O. Gökalp, A. Koçyiğit, and P. E. Eren, “A visual programming framework for distributed Internet of Things centric complex event processing,” *Comput. Electr. Eng.*, 2018.

[99] S. Ozarslan and P. E. Eren, “MobileCDP: A mobile framework for the consumer decision process,” *Inf. Syst. Front.*, pp. 1–22, 2015.

[100] I. Guyon and A. Elisseeff, “An introduction to feature extraction,” in *Feature extraction*, Springer, 2006, pp. 1–25.

[101] J.-D. Lee, C.-Y. Lin, and C.-H. Huang, "Novel features selection for gender classification," in *Mechatronics and Automation (ICMA), 2013 IEEE International Conference on*, 2013, pp. 785–790.

[102] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[103] V. Singh, V. Shokeen, and M. B. Singh, "Comparison of feature extraction algorithms for gender classification from face images," *Int. J. Eng. Res. Technol.*, vol. 2, no. 5, 2013.

[104] "Description of Static Face Images." [Online]. Available: http://what-when-how.com/face-recognition/local-representation-of-facial-features-face-image-modeling-and-representation-face-recognition-part-2/.

[105] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[106] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, no. Dec, pp. 113–141, 2000.

[107] K.-B. Duan, J. C. Rajapakse, and M. N. Nguyen, "One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification," in *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, 2007, pp. 47–56.

[108] X. Li, L. Wang, and E. Sung, "Multilabel SVM active learning for image classification," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*, 2004, vol. 4, pp. 2207–2210.

[109] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

[110] A. Majid, A. Khan, and A. M. Mirza, "Gender classification using discrete cosine transformation: a comparison of different classifiers," in *Multi Topic Conference, 2003. INMIC 2003. 7th International*, 2003, pp. 59–64.

[111] "File:KnnClassification.svg - Wikipedia." [Online]. Available: http://www.wikizero.org/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvRmlsZTpLbm5DbGFzc2lmaWNhdGlvbi5zdmc. [Accessed: 14-Jun-2017].

[112] H. H. K. Tin, "Perceived gender classification from face images," *Int. J. Mod. Educ. Comput. Sci.*, vol. 4, no. 1, p. 12, 2012.

[113] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural network design*, vol. 20. Pws Pub. Boston, 1996.

[114] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[115] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Neural networks, 1997., international conference on*, 1997, vol. 3, pp. 1930–1935.

[116] P. E. Gill, W. Murray, and M. H. Wright, “Practical optimization,” 1981.
[117] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.
[118] M. Riedmiller and H. Braun, “A direct adaptive method for faster backpropagation learning: The RPROP algorithm,” in *Neural Networks, 1993., IEEE International Conference on*, 1993, pp. 586–591.
[119] Y. LeCun *et al.*, “Backpropagation applied to handwritten zip code recognition,” *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
[120] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[121] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[122] Y. Zhang, W. Chan, and N. Jaitly, “Very deep convolutional networks for end-to-end speech recognition,” in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, 2017, pp. 4845–4849.
[123] C. Szegedy *et al.*, “Going deeper with convolutions,” 2015.
[124] Z. Lu, X. Jiang, and A. C. Kot, “Deep Coupled ResNet for Low-Resolution Face Recognition,” *IEEE Signal Process. Lett.*, 2018.
[125] D. Han, Q. Liu, and W. Fan, “A new image classification method using CNN transfer learning and web data augmentation,” *Expert Syst. Appl.*, vol. 95, pp. 43–56, 2018.
[126] Z. Tu *et al.*, “Multi-stream CNN: Learning representations based on human-related regions for action recognition,” *Pattern Recognit.*, vol. 79, pp. 32–43, 2018.
[127] G. Rogez and C. Schmid, “Image-based synthesis for deep 3d human pose estimation,” *Int. J. Comput. Vis.*, pp. 1–16, 2018.
[128] C. N. dos Santos, B. Xiang, and B. Zhou, “Text classification by ranking with convolutional neural networks.” Google Patents, 26-Oct-2017.
[129] A. Bhandare, M. Bhide, P. Gokhale, and R. Chandavarkar, “Applications of Convolutional Neural Networks,” *Int. J. Comput. Sci. Inf. Technol.*, pp. 2206–2215, 2016.
[130] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, “Identification of rice diseases using deep convolutional neural networks,” *Neurocomputing*, vol. 267, pp. 378–384, 2017.
[131] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
[132] A. Loomis, *Figure drawing for all it’s worth*. Viking Pr, 1943.
[133] Z. Zhang, “Microsoft kinect sensor and its effect,” *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, 2012.

[134] “Skeleton Position and Tracking State.” [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131025.aspx. [Accessed: 14-Jun-2017].

[135] C.-Y. Tsai, C.-H. Huang, and A.-H. Tsao, “Graphics processing unit-accelerated multi-resolution exhaustive search algorithm for real-time keypoint descriptor matching in high-dimensional spaces,” *IET Comput. Vis.*, vol. 10, no. 3, pp. 212–219, 2016.

[136] L. Bui, D. Tran, X. Huang, and G. Chetty, “Face gender classification based on active appearance model and fuzzy k-nearest neighbors,” in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, 2012, p. 1.

[137] S. Camalan and G. Sengul, “Gender prediction by using Local Binary Pattern and K Nearest Neighbor and Discriminant Analysis classifications,” in *2016 24th Signal Processing and Communication Application Conference, SIU 2016 - Proceedings*, 2016.

[138] J. Kacprzyk and W. Pedrycz, *Springer handbook of computational intelligence*. Springer, 2015.

[139] B. C. Munsell, A. Temlyakov, C. Qu, and S. Wang, “Person identification using full-body motion and anthropometric biometrics from kinect videos,” in *European Conference on Computer Vision*, 2012, pp. 91–100.

[140] E. Gianaria, M. Grangetto, M. Lucenteforte, and N. Balossino, “Human classification using gait features,” in *International Workshop on Biometric Authentication*, 2014, pp. 16–27.

[141] A. Jaswante, A. U. Khan, and B. Gour, “Back Propagation Neural Network Based Gender Classification Technique Based on Facial Features,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 14, no. 11, p. 91, 2014.

[142] S. Rudra *et al.*, “Gender classification system from offline survey data using neural networks,” in *Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual*, 2016, pp. 1–5.

[143] H. Tran, P. N. Pathirana, and A. Seneviratne, “Human Gender Recognition with Upper Body Gait Kinematics.”

[144] K.-B. Duan and S. S. Keerthi, “Which is the best multiclass SVM method? An empirical study,” in *International workshop on multiple classifier systems*, 2005, pp. 278–285.

[145] “Pretrained Convolutional Neural Networks - MATLAB & Simulink.” [Online]. Available: https://www.mathworks.com/help/nnet/ug/pretrained-convolutional-neural-networks.html. [Accessed: 26-May-2018].

[146] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248–255.

[147] S. Escalera *et al.*, “Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 1–9.

[148] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 10–15.
[149] "IMDB-WIKI - 500k+ face images with age and gender labels." [Online]. Available: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/. [Accessed: 25-May-2018].
[150] "The Images of Groups Dataset." [Online]. Available: http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html. [Accessed: 27-May-2018].
[151] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu, "Age and gender estimation based on wrinkle texture and color of facial images," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, vol. 1, pp. 405–408.
[152] A. Sandygulova, Y. Absattar, D. Doszhan, and G. I. Parisi, "Child-Centred Motion-Based Age and Gender Estimation with Neural Network Learning.," in *AAAI Workshop: Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, 2016.
[153] E. Malmi and I. Weber, "You Are What Apps You Use: Demographic Prediction Based on User's Apps.," in *ICWSM*, 2016, pp. 635–638.
[154] H. A. Schwartz *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLoS One*, vol. 8, no. 9, p. e73791, 2013.
[155] J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen, "Demographic prediction based on user's browsing behavior," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 151–160.
[156] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 34–42.
[157] "How old do I look?" [Online]. Available: https://how-old.net/#. [Accessed: 23-May-2018].
[158] "Online Age Detector." [Online]. Available: http://agedetector.tequnique.com/. [Accessed: 23-May-2018].
[159] "Kairos: Determining How Old You Are From Photos and Video." [Online]. Available: https://www.kairos.com/blog/determining-how-old-you-are-from-photos-and-video. [Accessed: 23-May-2018].
[160] V. Almeida, M. K. Dutta, C. M. Travieso, A. Singh, and J. B. Alonso, "Automatic age detection based on facial images," in *Communication Control and Intelligent Systems (CCIS), 2016 2nd International Conference on*, 2016, pp. 110–114.
[161] A. Schmidt, "Cloud-based ai for pervasive applications," *IEEE Pervasive Comput.*, vol. 15, no. 1, pp. 14–18, 2016.
[162] C. Henderson and E. Izquierdo, "Reflection Invariance: an important consideration of image orientation," *arXiv Prepr. arXiv1506.02432*, 2015.

[163] L. Quinn and M. Lech, “Multi-stage classification network for automatic age estimation from facial images,” in *Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on*, 2015, pp. 1–6.

[164] G. Wu, “The relation between age-related changes in neuromusculoskeletal system and dynamic postural responses to balance disturbance,” *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.*, vol. 53, no. 4, pp. M320–M326, 1998.

[165] Y. Ge, J. Lu, X. Feng, and D. Yang, “Body-based human age estimation at a distance,” in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1–4.

[166] M. H. Zaki and T. Sayed, “Using automated walking gait analysis for the identification of pedestrian attributes,” *Transp. Res. part C Emerg. Technol.*, vol. 48, pp. 16–36, 2014.

[167] A. Dantcheva and F. Brémond, “Gender estimation based on smile-dynamics,” *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 3, pp. 719–729, 2017.

[168] P. Bilinski, A. Dantcheva, and F. Brémond, “Can a smile reveal your gender?,” in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, 2016, pp. 1–6.

[169] A. Lanitis, “Age estimation based on head movements: A feasibility study,” in *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, 2010, pp. 1–6.

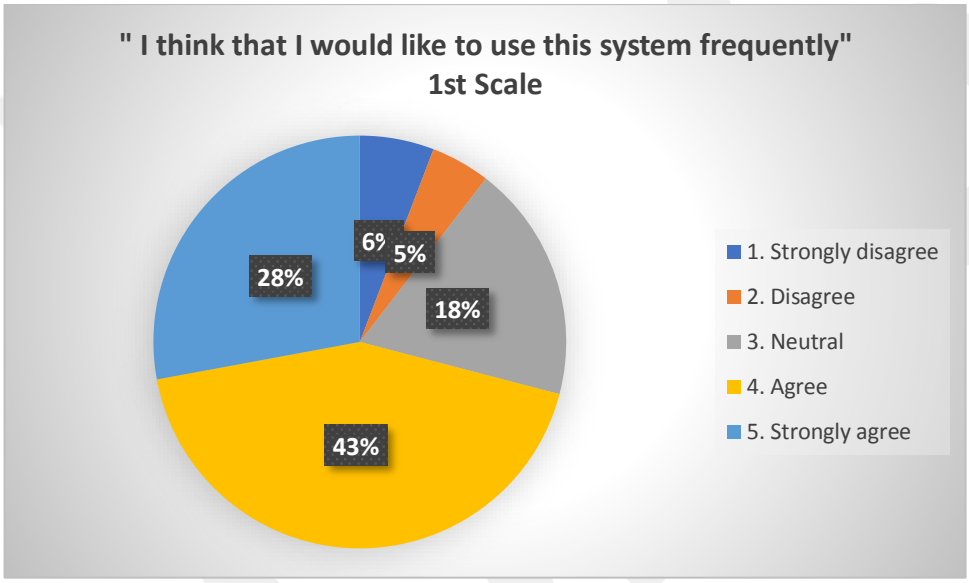
APPENDIX

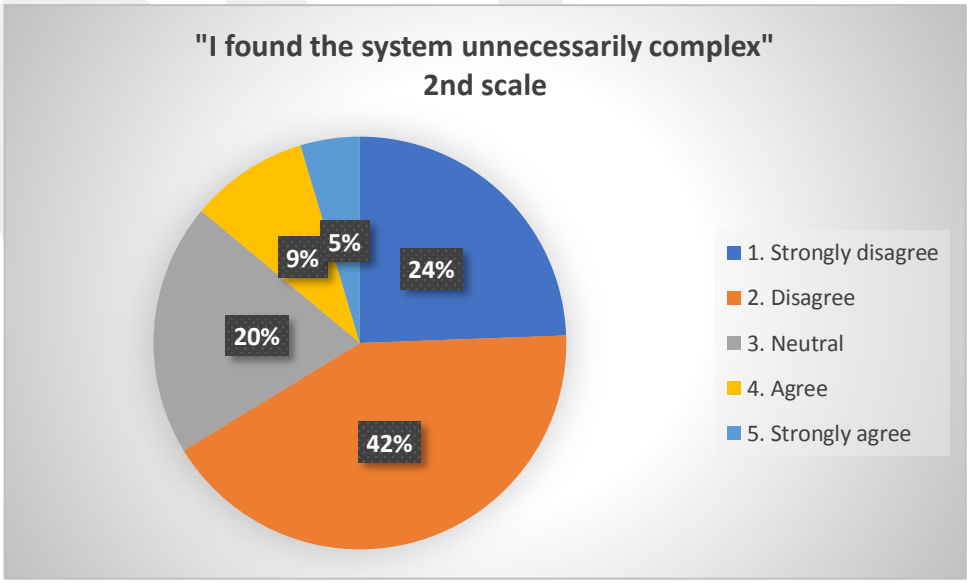
A. Participants Data of Accuracy Test

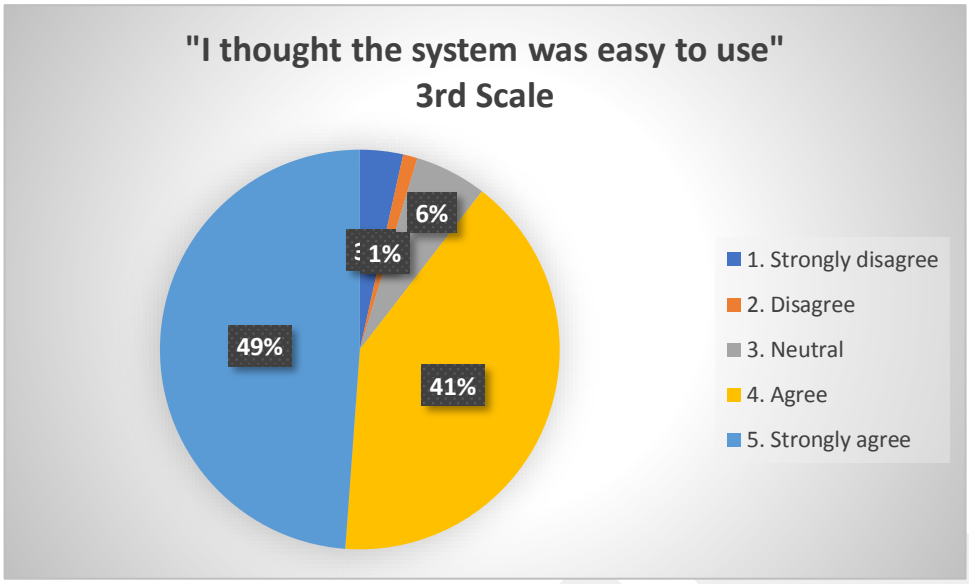
Participant	*Actual Age*	*Predicted Age*	*Actual Gender*	*Predicted Gender*
Person 1	24	20-36	Male	Male
Person 2	35	20-36	Male	Male
Person 3	25	3-8	Female	Female
Person 4	22	20-36	Male	Male
Person 5	22	3-8	Male	Male
Person 6	21	3-8	Female	Female
Person 7	24	9-12	Male	Male
Person 8	29	20-36	Male	Male
Person 9	24	3-8	Male	Male
Person 10	24	9-12	Male	Male
Person 11	24	20-36	Female	Male
Person 12	21	20-36	Male	Male
Person 13	23	20-36	Male	Male
Person 14	24	20-36	Male	Male
Person 15	21	9-12	Male	Male
Person 16	26	20-36	Male	Male
Person 17	24	13-19	Male	Male
Person 18	24	20-36	Male	Male
Person 19	25	9-12	Male	Male
Person 20	21	13-19	Male	Male
Person 21	24	20-36	Male	Male
Person 22	30	20-36	Male	Male
Person 23	28	20-36	Male	Male
Person 24	23	13-19	Male	Male
Person 25	22	20-36	Male	Male
Person 26	25	3-8	Male	Male
Person 27	24	20-36	Male	Male
Person 28	25	3-8	Female	Male
Person 29	50	20-36	Male	Male
Person 30	24	9-12	Male	Male
Person 31	23	20-36	Male	Male
Person 32	25	20-36	Male	Male
Person 33	24	20-36	Male	Male
Person 34	25	20-36	Female	Female
Person 35	23	20-36	Female	Male
Person 36	25	9-12	Male	Male
Person 37	24	20-36	Male	Male
Person 38	23	9-12	Male	Male
Person 39	21	20-36	Male	Male
Person 40	28	20-36	Male	Male

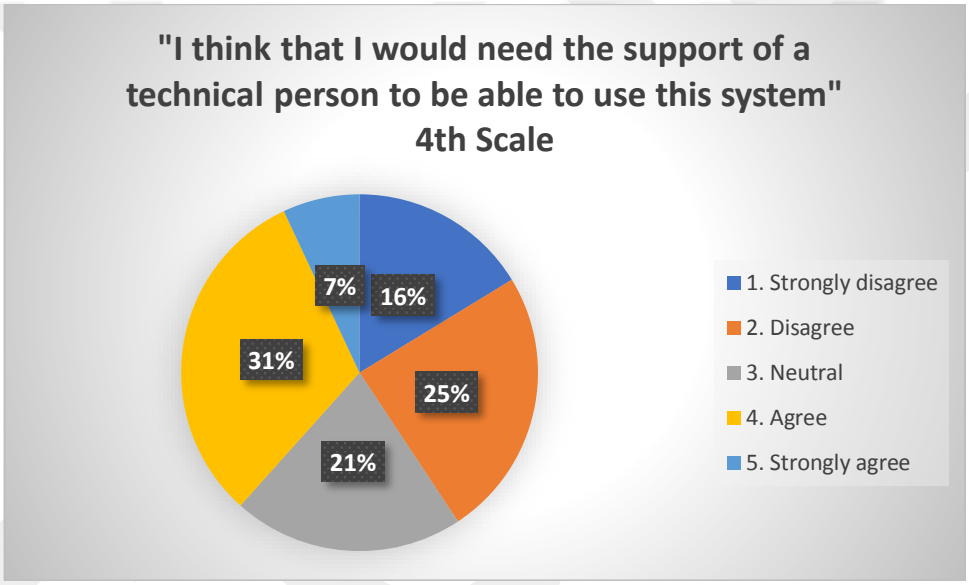
Person 41	23	13-19	Male	Male
Person 42	28	9-12	Female	Male
Person 43	29	20-36	Male	Male
Person 44	22	13-19	Female	Male
Person 45	29	13-19	Female	Male
Person 46	28	20-36	Male	Male
Person 47	29	20-36	Female	Male
Person 48	30	20-36	Male	Male
Person 49	24	20-36	Male	Male
Person 50	42	20-36	Male	Male

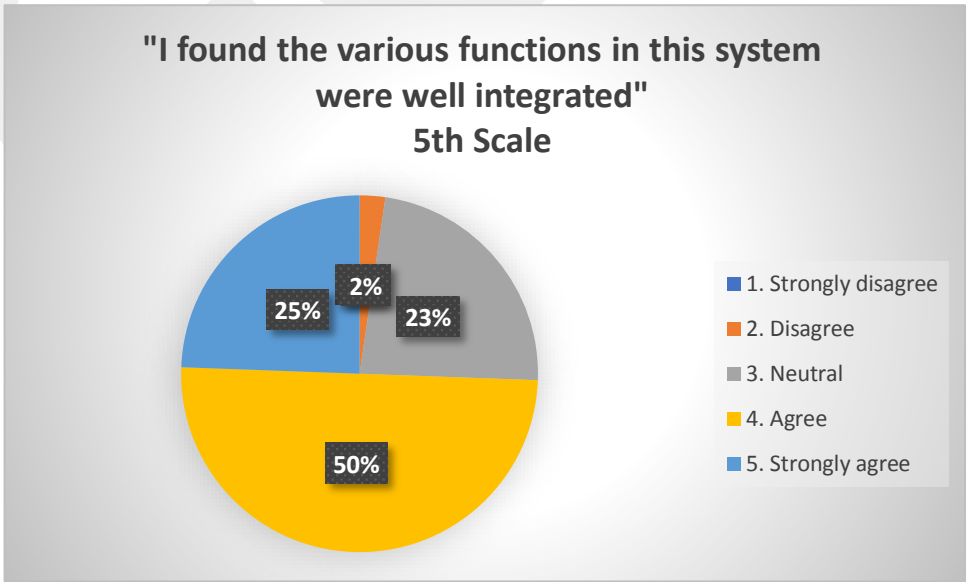
B. SUS Pie Chart Representations

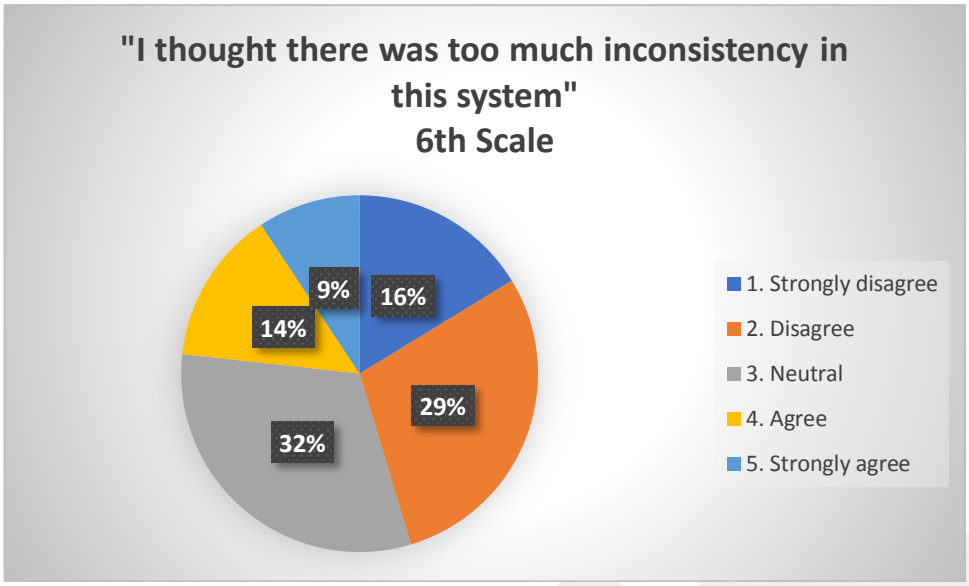


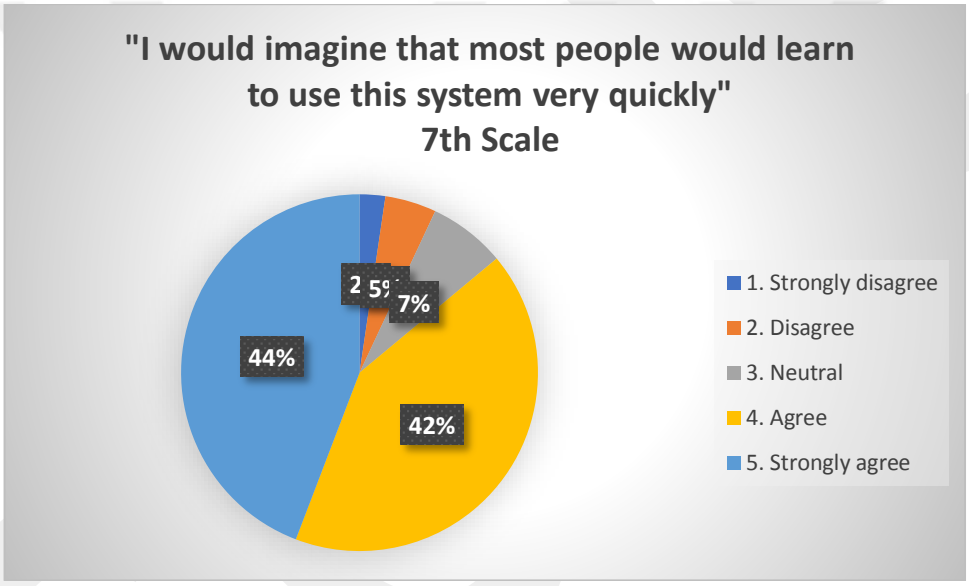


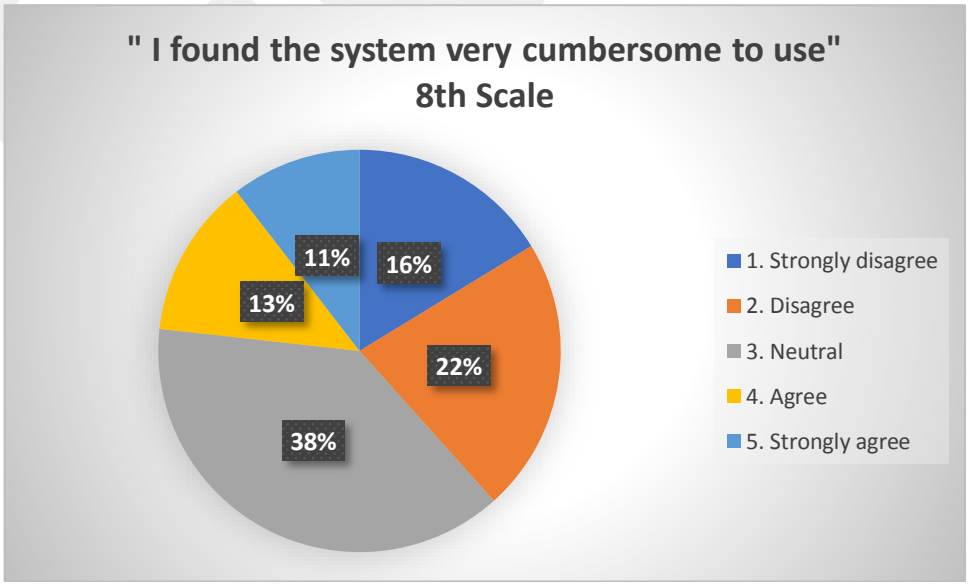

"I thought the system was easy to use"
3rd Scale
6%
1%
49%
41%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

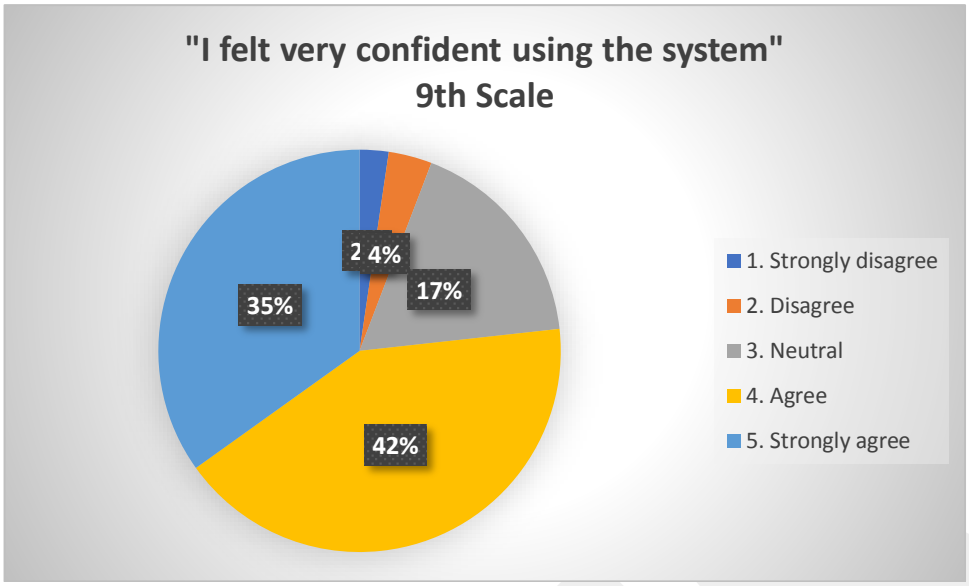

"I think that I would need the support of a technical person to be able to use this system"
4th Scale
7%
16%
31%
25%
21%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

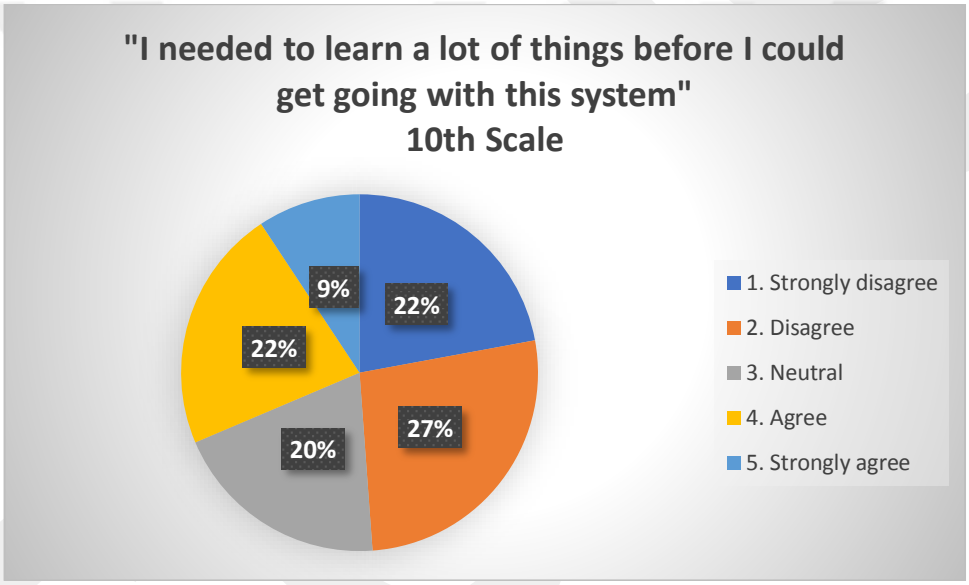

"I found the various functions in this system were well integrated"
5th Scale
25%
2%
23%
50%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree


"I thought there was too much inconsistency in this system"
6th Scale
16%
29%
32%
14%
9%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree


"I would imagine that most people would learn to use this system very quickly"
7th Scale
7%
44%
42%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree


" I found the system very cumbersome to use"
8th Scale
16%
22%
38%
13%
11%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree


"I felt very confident using the system"
9th Scale
2 4%
17%
35%
42%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree


"I needed to learn a lot of things before I could get going with this system"
10th Scale
9%
22%
22%
27%
20%
1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree