

Face Recognition Using Convolutional Neural Network (CNN)



By

Thomas Endalemaw Ahmed

A Thesis Submitted to

The Department of Computer Science and Engineering,

School of Electrical and Computing

**Presented in Partial Fulfillment of Master's Degree in Computer
Science and Engineering**

Office of Graduate Studies

Adama Science and Technology University

January, 2022

Adama, Ethiopia

Face Recognition Using Convolutional Neural Network (CNN)

By

Thomas Endalemaw Ahmed

Advisor:

Tilahun Melak Sitote (PhD)

A Thesis Submitted to

The Department of Computer Science and Engineering,

School of Electrical and Computing

**Presented in Partial Fulfillment of Master's Degree in Computer
Science and Engineering**

Office of Graduate Studies

Adama Science and Technology University

January, 2022

Adama, Ethiopia

APPROVAL SHEET

The Thesis entitled “**Face Recognition Using Convolutional Neural Network (CNN)**” and developed by **Thomas Endalemaw Ahmed**. Hear by certifying that the recommendation and Suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis

Tilahun Melak (Ph.D.)

Advisor

Signature

Date

We, the undersigned, members of the board of Examiners of the thesis by **Thomas Endalemaw Ahmed** Have read and evaluated the thesis entitled “**Face Recognition Using Convolutional Neural Network (CNN)**” and examined the candidate during the open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Masters of Science in Computer Science and Engineering.

Chairperson

Signature

Date

Internal Examiner

Signature

Date

External Examiner

Signature

Date

Final approval and acceptance of the thesis are contingent upon submission of its Final copy to the office of postgraduate Studies (**OPGS**) through the Department Graduate Council (**DGC**) and School Graduate Committee (**SGC**).

Department Head

Signature

Date

School Dean

Signature

Date

Legesse Lemeche Obsu (PH.D.)

Office of postgraduate studies. Dean

Signature

Date

DECLARATION

I hereby declare that this MSc Thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been properly acknowledged.

Thomas Endalemaw

Name Student

Signature

Date

RECOMMENDATION

I, the advisor of this thesis, hereby certify that we have read the revised version of the thesis entitled “**Face Recognition Using Convolutional Neural Network (CNN)**” prepared under my/our guidance by **Thomas Endalemaw Ahmed** Submitted in Partial Fulfillment of the requirement for the degree of master’s degree of Computer Science and Engineering

Therefore, I recommend submitting the revised of the thesis to the department following the applicable procedures.

Tilahun Melak (PH. D)

Advisor

Signature

Date

ACKNOWLEDGMENT

First and foremost, I'd like to thank God next to God I'd like to thank my adviser Tilahun Melak (Ph.D), for his vital effort in advising, checking, and providing knowledgeable, constructive, and valuable comments by the quality of his professional experience Finally, I love you so much, my Families Thank you for being with me all time. May God Grant you long life and Health. And. I would like to express my genuine thank to my friends for their supports.

TABLE OF CONTENTS

APPROVAL SHEET	III
DECLARATION.....	IV
RECOMMENDATION	V
ACKNOWLEDGMENT	VI
LIST OF FIGURES	X
LIST OF TABLES	XI
LIST OF ACRONYMS	XII
ABSTRACT	XIII
CHAPTER ONE: INTRODUCTION	1
1.1 Background of the Study	1
1.2 Motivation of the Study	3
1.3 Statement of the problems.....	3
1.4 Objectives of the Study	4
1.4.1 General Objective	4
1.4.2 Specific Objective.....	4
1.5 Significance of the Study	4
1.6 Methodologies.....	5
1.6.1 Tools	5
1.7 Scope and Limitations	5
1.8 Application of Results.....	6
1.9 Organization of the Thesis	6
CHAPTER TWO: LITERATURE REVIEW AND RELATED WORK.....	7
2.1. Introduction.....	7
2.3 Human Face Recognition	9
2.4 Machine Recognition of Faces	10
2.5 Image Processing.....	10
2.6 Face Detection	11
2.7 Face Identification	11
2.8 Face Recognition Approaches.....	12
2.8.1 Holistic approach to face recognition.....	12
2.8.2 Statistical approaches to face recognition.....	12
2.8.3 Hybrid approaches for face recognition	12

2.8.4	Feature based approaches for face recognition.....	13
2.9	Mathematical Description in Face Recognition and Detection.....	13
2.9.1	Based on Mathematical Operation.....	13
2.9.1.1	Histogram based operation	13
2.10	Related Work	14
CHAPTER THREE: RESEARCH METHODOLOGY.....		19
3.1	Overview	19
3.2	Data collection	19
3.3	Development Tools.....	20
3.3.1	Software Tools.....	20
3.3.2	Hardware Tool	24
3.4	Image Acquisition	24
3.5	Image Data Pre-processing	24
3.6	Image Data Augmentation	24
3.8	Performance Evaluation Metrics.....	26
CHAPTER FOUR: RESEARCH DESIGN.....		28
4.1.	Introduction.....	28
4.2	Model Selection	28
4.3	Model Design.....	28
4.3.1	Input Layer.....	30
4.3.2	Convolutional Layer (Convlayer).....	30
4.3.3	Pooling Layer	30
4.3.4	Fully Connected (FC) Layer	31
4.3.5	Output Layer.....	31
4.6	Image conversion to gray-scale.....	35
4.7	Algorithms for Optimization.....	37
4.8	Activation function.....	37
4.8.1	Epochs	37
4.8.2	Batch size	38
4.8.3	Learning rate.....	38
4.9	Hyper-parameters for TomNet Model.....	38
4.10	Histogram Equalizer.....	39
CHAPTER FIVE: IMPLEMENTATION		40

5.1 Proposed System	40
5.2 Data Preparation.....	40
5.3 Image Preprocessing.....	42
5.4 Haar features.....	47
5.5 Description of Module	47
5.5.1 Correct Face Position and Orientation	48
5.5.2 Face Recognition	49
5.5.3 Histogram Matching.....	49
CHAPTER SIX: RESULT, DISCUSSION AND EVALUATION	51
6.1 Discussion.....	51
6.2 Result.....	51
6.3 Result of Experiment done on RGB data images.....	52
6.3.1 Experiment result using output Activation Function	52
6.3.2 Experiment result using various Learning Rates.....	52
6.3.2.1 Activation Function.....	53
6.4 Evaluation result	57
CHAPTER SEVEN: CONCLUSION AND FUTURE WORK	60
7.1 Conclusion	60
7.2 Future Work.....	60
REFERENCE	61
APPENDIXES	65

LIST OF FIGURES

Figure 1 A general CNN architecture	7
Figure 2 Diagram of OpenCV.....	22
Figure 3 PCA vs LDA.....	23
Figure 4 Face Recognition and Detection Diagram.....	29
Figure 5 Max Pooling and Average Pooling.....	30
Figure 6 Full Connected layer.....	31
Figure 7 Input structure of Dataset	34
Figure 8 Example of the training image Database.....	35
Figure 9 Sample original image.....	41
Figure 10 Augmented Images	41
Figure 11 Sample code for data augmentation	42
Figure 12 Sample code for TomNet model.....	44
Figure 13 Haar Like Features in Viola Jones.....	48
Figure 14 Artificial Neural Networks	49
Figure 15 Confusion Matrix for CNN of images.....	54
Figure 16 Training and Validation Accuracy Analysis.....	52
Figure 17 Training and Validation Loss Analysis	55
Figure 18 Age Histogram.....	55
Figure 19 Ethnicity Histogram.....	56
Figure 20 Gender Histogram	56
Figure 21 Accuracy Graph of Train and Validation of images	58
Figure 22 Loss Graph of Train and Validation of images	59

LIST OF TABLES

Table 1 comparison of different face databases.....	16
Table 2 A summary of the research results depending on different methods.....	17
Table 3 Confusion matrix	26
Table 4 TomNet model summary	31
Table 5 Summary of hyper-parameters used during model training	38
Table 6 Sample Code for TomNet_model	45
Table 7 Experiment result using learning rate	52
Table 8 Experiment of the TomNet model using different learning rate.....	53
Table 9 Experiment result using different activation function on images	53
Table 10 Evaluation summary for the TomNet model	57

LIST OF ACRONYMS

CNN	Convolutional Neural Network
CPU	Central Processing Unit
RAM	Random Access Memory
SVM	Support Vector Machine
ANN	Artificial Neural Network
Adam	Adaptive learning rate optimization algorithm
ConvNets	Convolutional neural network
Conv	Convolution
FN	False Negative
FP	False Positives
GH _z	Giga hertz
GB	Gigabytes
TN	True Negative
TP	True Positives
Val_accuracy	Validation accuracy
Val_loss	Validation loss
RELU	Rectified linear unit
RMSE	Root Mean Square Error

ABSTRACT

Face recognition is one of the most important applications in video surveillance and computer vision. However, the conventional algorithms of face recognition are susceptible to multiple conditions such as lighting occlusion viewing angle or camera rotation. Therefore face recognition based on deep learning can greatly improve the recognition speed and compatible external interface. In this thesis we use convolutional neural networks (ConvNets) for face recognition the neural networks have the merits of end-to-end sparse connection and weight sharing.

The purpose of this thesis is to identify the age, gender and ethnicity of different people based on location of the detected box of a face. Then we can obtain recognition result with different methods.

Image processing is the wide area of studies and need highly integrated software and hardware tool. Currently a number of researchers are used with python integrated with OpenCV, Keras and Tensorflow and this proposed study is works in the same. The proposed research used convolutional neural network algorithms because it has speed when compared to other algorithms. The proposed algorithms used 19108 individual images with different features. Datasets are split in to training and other for Testing. The efficiency of the model in this research experiment is measured by deep learning approaches the confusion matrix is used to assess model performance. The TomNet model is tested in different epochs and learning rates to generate an efficient model. The result of the experiment is 98.88% training accuracy with 98.56% validation accuracy.

The use of neural network for face recognition improves the speed of recognition. The contribution of this thesis is: can identify a human face including rotation and position. The confidence of human face recognition is mainly affected by the proportion of face on the screen.

Keyword: Deep Learning, Convolutional Neural Network, Face Recognition, Computer Vision, Face Detection

THIS PAGE IS HELD FOR MY MOTHER.

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

A facial recognition system is a technology capable of identifying or verifying a person from a digital image or video frame from the database. It is typically used as access control in a security system and can be compared to other bio-metrics such as fingerprint or eye iris recognition system.

Face recognition is an application to naturally distinguish an individual in the wake of contrasting and a database of pictures. A test picture is taken with the assistance of a camera from a still source or video outline, and in the wake of representing different conceivable face appearance varieties, it is contrasted and the database pictures for the character. [1]

Facial Recognition is one of the most emerging and much used application of image processing which has gained much importance and because of vast usage of biometric system during past years. Biometric based techniques are emerged to be the replacement of recognizing individual instead of using Passwords, keys, PINs, cards etc. It is hard to remember them, and they can be stolen, or one can forget them, and they can be easily lost. However, there is no such headache in use of biometric based techniques. Biometric based identification includes Facial Recognition, Finger or thumb prints, Retina or voice-based authentication. Face recognition have advantages over other methods, in other methods user's action is also required like placing thumb on biometric fingerprint verification machine or standing in fixed position for retina identification. Using voice recognition in crowd can be difficult task due to background noise. Facial recognition is the most economic and easy solution to these problems. One more reason is that Facial Recognition has been reliable in many sectors in various ways e.g., commercial, forensic, security etc. [2]

The goal of face detection is to determine whether or not there are any faces in the image and if present, return the image location and extend of each face. Practically face detection algorithm will aim at generating bounding boxes (often rectangular or elliptical) around all the face in the image and only around faces. [3]

A formal method of classifying faces was first proposed by Francis Galton in 1888 [4,5]. During 1980's work on FR remained largely dormant. However, during 1990's the research interest in FN has grown significantly as a result of the following facts:

1. The increase in emphasis on civilian/commercial research projects
2. The emergence of CNN classifiers with emphasis on real time computation and adaptation
3. The availability of real time hardware
4. The increasing need for surveillance related applications due to drug trafficking, terrorist activities, financial sectors etc.

In terms of algorithms, there are sharing parameters between the convolution layer and the convolution layer of CNN. The advantage of this is that the memory requirements are reduced, and the number of parameters to be trained is correspondingly reduced. The performance of the algorithm is therefore improved. At the same time, in other machine learning algorithms, the pictures need us to perform preprocessing or feature extraction. However, we rarely need to do these operations when using CNN for image processing. [6]

The idea of Face detection and Recognition is generated after Emerging of Computer Vision. Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos [7]. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. What distinguished computer vision from the prevalent field of digital image processing at that time was a desire to extract three-dimensional structure from images with the goal of achieving full scene understanding. Studies in the 1970s formed the early foundations for many of the computer vision algorithms that exist today, including extraction of edges from images, labeling of lines, no polyhedral and polyhedral modeling, and representation of objects as interconnections of smaller structures, optical flow, and motion estimation. The next decade saw studies based on more rigorous mathematical analysis and quantitative aspects of computer vision. [8]

1.2 Motivation of the Study

Currently in the era of digital human interest is increase with advancement of Technology through any fields of studies. Especially after generation and emerging of Computer Vision, image related research is highly motivated and hot research area. The development of image related research becomes the pioneer of Identification and detection. This identification and detection is also the fundamental of Recognition. The importance of Detection and Recognition become widely known and applicable after computer vision is used for face detection and Recognition. But the emergence of this technology does not mean that we are satisfied due to quality and per needed detected and recognized face. Even though many face detection and recognition research and project exists over the globe it faced the challenge on quality, dynamicity, accuracy and reliability. Therefore, it is needed to contribute algorithms for these interest fields of study.

1.3 Statement of the problems

Face detection and recognition are important for many reasons and thus most computer vision algorithms are dedicated to the task of facial feature analysis and recognition. But it is open area for research in terms of accuracy and reliability [9]. For most face detection algorithms faces are not tracked but the detection algorithm is run on each frame from a video or streaming camera. The more recent approaches to face detection and recognition are based on detecting facial landmarks i.e. nose, mouth, right eye and left eye. It can have done by training several Convolutional Neural Networks (CNN) for detecting different facial landmarks and combine the results in the output network [10].

As compared to different biometric techniques appearance identification may not be the most reliable and effective. Quality measurements are very important in facial recognition system as huge points of variation are likely in face pictures. Elements e.g. light reflection, posture, and sound during face action may affect the performance of facial recognition schemes. Among all biometric schemes facial recognition has the highest false approval and act rates so doubts have been raised on the strength of face identification packages in instances of line and airport security. The main question and problem of statement for this research are:

RQ1. What kind of technology can achieve face recognition?

RQ2. How to develop a face recognition model based using an image analysis technique?

RQ3. What is the feature taken with the algorithm when face is recognition?

RQ4. What are the significant features that can be used to identify face recognition?

1.4 Objectives of the Study

1.4.1 General Objective

The general objective of this study is to develop a computer-based face recognition model using deep learning approaches.

1.4.2 Specific Objective

To achieve the general objectives of this research, the research is consisting of the following specific objectives:

- ❖ Perform a comprehensive literature review on face recognition using the CNN method
- ❖ Explore the available face recognition tools and techniques (algorithms) that are more convenient to carry out the research.
- ❖ Design a convolutional neural network model to detect faces
- ❖ Train the proposed model by using the dataset
- ❖ Evaluate the performance of face recognition using CNN algorithm and report the finding of this research work.
- ❖ Design a face recognition system to perform
 - Image preprocessing
 - Feature extraction using CNN
 - Face recognition
- ❖ Draw conclusion and suggest recommendation for further research on the area

1.5 Significance of the Study

The study provides different benefits for different financial banking sectors and other researchers. After the evaluations of the result, it can be applied in different areas:

- Governmental and private banking industries were benefit from the study in other way customers also beneficiaries of this study.

- The study opens a door for different researchers in Ethiopia who have the interest to work on financial securities.
- The model also can be applied to other Face recognition and Detection which can be detected based on RGB and Gray-scale images.

1.6 Methodologies

The research done on methodology related publication research, journal, conference paper, and surveys from different digital library and could include both present and historical information.

Simple testing of this program will also need to occur so that we could evaluate its effectiveness. Several sample images will be taken of several individuals to be used as test cases one each for account images and several each for lie images each of which would vary pose, lighting condition and expressions.

1.6.1 Tools

Selection of the appropriate tool has play for the success of the research. The Anaconda 3 python 3 are the high-level programming tools used to write code that used check the input human faces. Face is the unique identification for any human being; OpenCV comes with a trainer as well detector inputted dataset images. Yale face database is used to get the input faces. I have used the following digital libraries to explore my papers and extract relevant information out of these common and important digital libraries which are:

1. IEEE Explore.
2. Science Direct.

1.7 Scope and Limitations

The scope of this research is to develop a face recognition system using frontal face image. Face detection is out of the scope of this thesis. The major limitation of this research is it have small number of local faces. This is because of unavailability of face image databases which was designed for such purpose in Ethiopia. Due to the restriction of time in the completion of this research, it was difficult to construct a larger face database. Also, another limitation of the propose system is it the inputs are all frontal face images.

1.8 Application of Results

Biometrics has been widely used in forensics applications such as criminal identification and prison security.

Face Detection and Recognition are also used as a strong potential to be widely adopted in social application such as e-Banking, e-Commerce, and access control. Due to the huge increase in the use of electronic transactions, e-Banking and e-Commerce are becoming the most important emerging application of biometrics. Credit card and smart card security, ATM security, check cashing and fund transfers; online transactions and web access are some of the examples of the applications.

Face Detection and Recognition are use either in government and non-government organization as its necessity. It can also use in law enforcement. Face recognition minimizing victim trauma verifying Identify for court records. Since it is system face it is unbiased. So, it is used as tool of justice. It gives all report about victim and then simple to give judge.

The other Application of Face Detection and Recognition is for Security Purpose, Access control, comparing surveillance images to know terrorist. Thus, Face Detection and Recognition can be used as tool of Counterterrorism. Face recognition can also work as a means of access control to ensure that only authorized individuals get into facilities like labs, boardrooms, bank vaults, training centers for athletes and other sensitive location.

1.9 Organization of the Thesis

Chapter one introduces the background concept and idea about this research paper. Chapter two presents the review of literature that discusses CNN and related works. Chapter three of this research paper deals with methodologies, methods and tools used in conducting this research. Chapter four is about the implementation of the system as well as the results of the research experiment. Chapter five covers discussions, analysis, and evaluation of the result whereas chapter six generalizes the research using conclusions and recommendations. References, and programming scripts are also included as appendixes.

CHAPTER TWO: LITERATURE REVIEW AND RELATED WORK

2.1. Introduction

This chapter focuses on the Face Recognition and Detection methods and techniques. using deep learning techniques and approaches for real world application. In face recognition, every system starts with inputting a human face for the machine. After recognizing the faces system process the activity given by the user. In this chapter, explain the science of face recognition base on deep learning.

2.2 Convolutional Neural Network (CNN)

A convolutional neural network is preferred as a deep learning method in this study. CNN, which can easily identify and classify objects with minimal pre-processing, is successful in analyzing the visual image and can easily separate the required features with its multi-layered structure. It consists of four main layers: - convolutional layer, pooling layer, activation function layer, and fully connected layer. Fig 2.1 shows a general CNN architecture. Layers of CNN are: - Convolution layer, Activation layer (e.g., using ReLu), pooling layer, Dropout Layer, fully connected Layer.

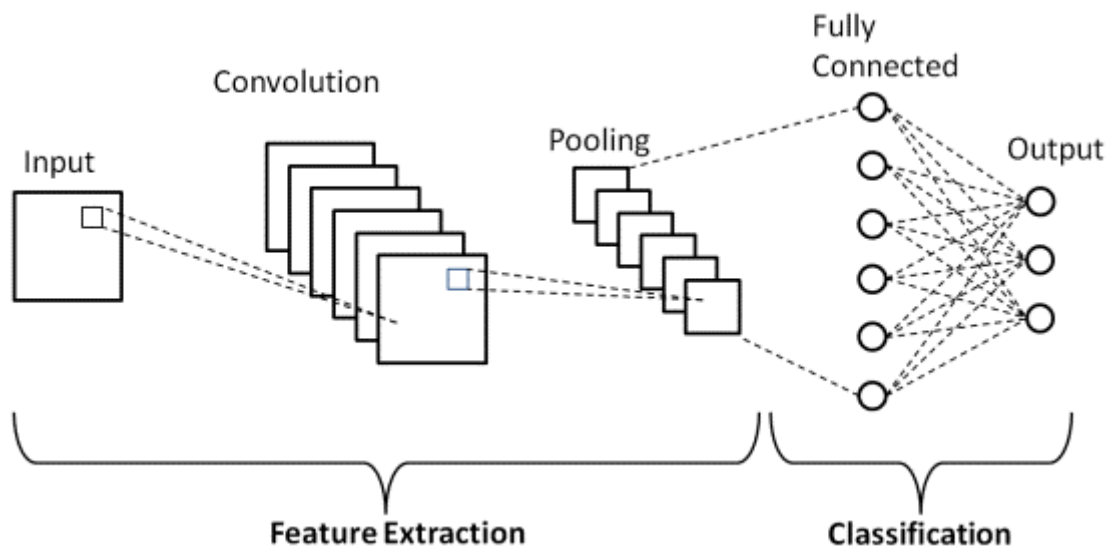


Figure 1 A general CNN architecture

(a) Convolution Layer A filter or kernel is used to run over the image in fixed gap intervals called strides. Selecting the size of stride is crucial to achieving the desired results. During

running the filter over the image, the dot product of filter with part of the image on which filter lies is calculated. Then the sum of all values of the product matrix is copied to the corresponding position in the convolved feature map matrix. Thus, we get a reduced dimension feature map of the image. Filters may be of many kinds, where each filter is used to extract a different kind of feature from the image.

(b) Parameters which helps in adjusting CNN's performance Stride –this defines the number of pixels by which we have to move our filter over the image so that we can focus on a new set of pixels while doing convolution. Stride's value ranges from 1 to 3 depending upon the amount of loss that we can be accommodated during convolution. The amount of loss in the image increases with the increasing value of stride.

(c) Dimension Reduction in CNN While convolution operation, the dimension of the input image is reduced after a filter matrix is convoluted over the input image. It also increases the depth of the image as this helps in identifying the regions in the image. E.g., a 9×9 RGB image with 3, 3×3 kernels at stride 1(per channel), will produce an RGB feature of $7 \times 7 \times 3$. Here 3rd dimension represents the depth of the convolved image.

(d) Pooling Layer This layer operates a small kernel on the image at a fixed stride. It is used to pick the pixel with the highest intensity and discard other pixels. The resultant matrix will be a reduced dimensional matrix of the feature image. This helps in reducing the unnecessary sparse cells of the image which are of no use in classification. Max pooling helps reduce the dimensionality of the network (or image), but this may cause some information loss. The concept behind these is that adjacent or nearby pixels can be approximated by the maximum information-carrying pixel.

(e) Activation Layer This layer mostly uses ReLu, softmax, sigmoid as activation. Function ReLu is a function that is used to set all negative values to zero and keeps positive value as it is. This step is followed by the convolution and pooling layer. In numerous fields such as computer vision, they are becoming the state of art achieving near human or better performance. They sound interested but designing a CNN is a herculean task in itself. Till now there is no fixed formula for the design of CNN. Many researchers have come up with general suggestions such as. But they don't always hold, as the task-dependent as importantly on data as it does on the algorithm. CNN exploits the spatial hierarchical feature of data, extracting features, and helps

classify them into different classes. This has led to the development of a stream of data augmentation and pre-processing to increase the data, as more data allows for the chance of better training and avoiding over fitting. This helps build models that are more robust to new samples as we try to make it more generalized to noise at the training phase.

(f) Dropout Layer The dropout layer is usually applied after the layer containing neurons in the fully connected network. The dropout layer is a regularization layer. It randomly drops out the weights of some neurons in every iteration so that the other neurons get more weightage in that iteration. Usually, people take dropouts of 20-50%. By using the dropout layer, normally the performance of a network increases 1 to 2%.

(g) Fully Connected Layer The structure from top to down usually forms a pyramid structure, the number of parameters in these layers keep on converging until they finally reach the number of desired classes. Increasing the number of hidden units in the layer can increase the learning ability of the network, but there is a saturation of the increase in the accuracy of the network. There is no formulation of the units you choose as it is a hit and trial usually. These also pile up with the fully connected subset of the network. Most of the network in research usually performs well with the number of units in multiples of 64. Two to three layers' networks are good if enough patterns are being passed to the network after flattening of the convolutional layers.

2.3 Human Face Recognition

Like almost everything associated with the human body the brain, perceptive abilities, cognition and consciousness, face recognition in humans is wonder. We are not even close to understand of how we manage to do it. What is known is the temporal lobe in the brain is partly responsible for this ability. Damaged to the temporal lobe can result in the condition in which the patient can lose the ability to recognize face. This is this specific condition where an otherwise normal person who suffered some damage to a specific region in the temporal lobe loses the ability to recognize faces is called prosopagnosia. It is a very interesting condition when it occurs as the perception of faces remains normal (vision pathways and perception is fine) and the person can recognize people by their voice but not by faces [16, 17].

When building artificial face recognition systems, scientists try to understand the architecture of human face recognition system. The methodology of human face recognition system may be useful to understand the basic system. It's hard for human to remember many faces due to

storage limitations. A key potential advantage of a machine system is its memory capacity [16], whereas for a human face recognition system the important feature is its parallel processing capacity.

2.4 Machine Recognition of Faces

Face recognition from image is a sub area of the general object recognition problem. Application is law enforcement for mug shot identification, verification for personal identification such as driver licenses and credit cards, gateways to limited access areas, surveillance of crowd behavior are all potential applications of a successful face recognition system [18].

2.5 Image Processing

Digital image processing is a branch of computer vision in which images are converted into an array of integers or pixels or processing a digital image by using a digital computer so that we can get an enhanced image which we can extract useful information from it. Image processing is a technique to enhance raw images received from camera placed on satellites space probes and aircrafts or pictures taken in normal day to day life for various applications. This field of image processing significantly improved in recent times and extended to various field of science and technology. The image processing mainly deals with image acquisition, image enhancement, image segmentation, feature extraction, image classification etc.

In image preprocessing image data recorded by camera/sensors on a satellite restrain errors related to geometry and brightness values of the pixels. These errors are corrected using appropriate mathematical models which are either definite or statistical models. Image enhancement is the modification of image by changing the pixel brightness values to improve its visual impact. Image enhancement involves a collection of techniques that are used to improve the visual appearance of an image or to convert the image to a form which is better suited for human or machine interpretation.

Sometime images obtained from satellites and conventional and digital cameras lack in contrast and brightness because of the limitation of imaging sub systems and illumination condition while capturing image. Image may have different types of noise. In image enhancement the goal is to accentuate certain image features for subsequent analysis or for image display [28].

2.6 Face Detection

Detection is the process of detecting or noticing something in the real world. Also, detection in this study is used to refer to detecting the human faces and test whether exists or not. The face detection process is to determine whether or not there is any face in the image and if present, return the image location and extent of each face. The challenges associated with the face detection process can be attributed to the following factors:

Haar-like features are evaluated through the use of a new image representation that generates a large set of features and uses the boosting algorithm AdaBoost to reduce degenerative tree of the boosted classifiers for robust and fast interferences only simple rectangular Haarlike features are used that provides a number of benefits like sort of ad-hoc domain knowledge is implied as well as a speed increase over pixel based systems, suggestive to Haar basis functions equivalent to intensity difference readings are quite easy to compute. Implementation of a system that used such features would provide a feature set that was far too large, hence the feature set must be only restricted to a small number of critical features which is achieved by boosting algorithm, Adaboost [3].

- **Occlusion:** Face may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Facial Expression:** The appearances of faces are directly affected by a person's facial expression.
- **Pose:** The images of a face vary due to the relative camera-face pose such as frontal, 45-degree profile, upside down etc.
- **Illumination:** when a face image is formed factors such as lighting condition and camera characteristics also an affect the appearance of a face.

2.7 Face Identification

The identification test works from the assumption that all faces in the test are known persons. The percentage of correct identification is then reported as the Correct Identification Rate or the percentage of false identification is reported as the False Identification Rate.

2.8 Face Recognition Approaches

A wide variety of approaches to machine recognition of faces has been published in the literature. In terms of the sensing modality a system can take 2D intensity images, color images, infrared images, and 3D images. In terms of viewing of angle, a system may be designed for frontal views, profile views, general views, or a combination of them. In terms of computational tools, a system can have used programmed knowledge rules, statistical decision rules; neural network rules etc. in general there are different approaches to face recognition problems.

2.8.1 Holistic approach to face recognition

Holistic face recognition utilizes global information from faces to perform face recognition. The global information for faces is fundamentally represented by a small number of features which are directly derived from the pixel information of face images [35]. This approach emphasizes with the global features rather than specific part of the face [34]. Then statistical data reduction techniques like PCA and LDA were employed to deal with this curse of dimensional.

In the holistic approaches the PCA usually give high similarities indiscriminately for two images from a single person or from two different persons and the LDA is also complex as there is a lot of within class variation due to differing facial expressions head orientations, lighting condition, distance from the machine, etc.

2.8.2 Statistical approaches to face recognition

Statistical methods include template matching based system where the training and test mages are matched by measuring the correlation between them. Moreover, statistical methods include the projection-based methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and others algorithms. In fact, projection-based system came out due to the shortcomings of the straightforward template matching based approaches.

2.8.3 Hybrid approaches for face recognition

The hybrid approach has a special status among face recognition systems as they combine different recognition approaches in either serial or parallel to overcome the shortcomings of individual methods. In classification mode the trained PCA BPNN and ICA BPNN are fed with new face images.

2.8.4 Feature based approaches for face recognition

The face can be recognized by extracting the relative position and other parameters of distinctive features such as eyes, mouth, nose, and chin. The algorithm described the overall geometrical configuration of face features by a vector of numerical data representing position and size of main facial features. Cagnoni and Poggi (1999) suggested a feature-based approach instead of a holistic approach to face recognition. They applied the eignface method to sub-images such as eye, nose, and mouth.

They also applied a rotation correction to the faces in order to obtain better results. The performance of PCA and ICA compared on face images. They argue that ICA encodes face images with statistically independent variables which are not necessarily associated with the orthogonal eigenvector. [21]

2.9 Mathematical Description in Face Recognition and Detection

2.9.1 Based on Mathematical Operation

2.9.1.1 Histogram based operation

An important class of point operations is based upon the manipulation of an image histogram or a region histogram.

A. Contrast stretching

Image is scanned in such a way that the resulting brightness values do not make full use of the available dynamic range. This can be easily observed in the histogram of the brightness. If the image is intended to go from brightness 0 to brightness $2^q - 1$ then one generally maps the 0% value.

$$b[m,n] = [2^q - 1]. A[m,n] - \frac{\min}{\max - \min} \text{-----} 2.1$$

B. Histogram Equalization

The most common histogram normalization technique is histogram equalization where one attempts to change the histogram through the use of a function $b=f(a)$ into a histogram that is constant for all brightness values. This can be especially useful when the image has been acquired under different circumstances.

C. Image Acquisition

Image acquisition process totally depends on the hardware system which may a sensor that is again a hardware device. A sensor converts light into electrical charges. The sensor inside a camera measures the reflected energy by the scene being imaged. The image sensors employed by most digital cameras is a charge coupled devices.

2.10 Related Work

Face recognition is one of the types of bio-metrics technology. Now a day's bio-metrics technology is become more secure and cheaper than the others technology. The financial institutes like banks and other organization use this technology to secure their system. To provide protection for executing transaction in this manner, financial bodies can also offer more and more extra services at the ATM which can make more profile and slash down on the coast of services [4]. There are many types of bio-metric technology scanning: retina, scan, face recognition, vein geometry, fingerprint verification etc.

- **Face Recognition-** this is one of the most flexible methods as it can be done without the person being aware that they are being scanned
- **Fingerprint Verification-** the fingerprint of any person remains the same throughout the life and no two fingerprints are ever same.
- **Scanning of Retina-** the pattern of the blood vessel at the back of every eye is absolutely unique and is never changing
- **Iris Scanning-** this is also very difficult to reproduce and stays the same with your entire lifetime.

When the ATMs use face recognition and performance is tested through the database. These databases have some limitation. The face database contains face images with a wide variety of poses, illumination angels, gestures, face occlusion, and illuminate colors, these images have not been adequately annotated, thus limiting their usefulness for evaluating the relative performance of face detection algorithms [11].

Weihang Wang et.al [11] the authors are presenting an efficient method for face recognition. this paper proposed a layer-by-layer training method of a deep convolutional neural network to help the convolutional neural network to converge and proposed a sample transformation method to avoid over-fitting. This method had reached an accuracy of 91% on ORL face database. To solve

the face recognition problem, this paper proposed a SIAMESE convolutional neural network which was trained on different parts and scales of a face and concatenated the face representation. The face recognition algorithm had reached an accuracy of 91% on ORL and 81% on LFW face database.

Sharma S et.al [12] the authors address the problems by pose variation and intensity variation. Also, the proposed paper investigates other discriminative methods for decreasing the tradeoff between accuracy and false acceptance rate. The paper describes the process involved in the face recognition like face alignment and feature extraction. The paper also emphasizes the importance of the face alignment; thus, the accuracy and False Acceptance Rate (FAR) is observed by using proposed technique. The computational analysis shows the better performance than other state-of-art approaches. The work has been done on Face Recognition Grand challenge (FRGC) dataset and giving accuracy of 96% with FAR of 0.1.

Ni Kadek Ayu Wirdiani et.al [13] The KNearest Neighbor algorithm uses neighborhood classification as the predictive value of a good instance value. K-NN includes an instance-based learning group. This paper developed face identification using Principal Component Analysis (PCA). the preprocessing method that used in this research is contrast stretching, grayscale and haar cascade segmentation. This research is registered 30 people, each person had 3 images used for training and 2 images used for testing. The result obtained from several test of k value gives the best accuracy is 81% with $k=1$.

To improve the discriminative and generalization ability of lightweight network for face recognition, we propose an efficient variable group convolutional network called VarGFaceNet. Variable group convolution is introduced by VarGNet to solve the conflict between small computational cost and the unbalance of computational intensity inside a block. We employ variable group convolution to design our network which can support large scale face identification while reduce computational cost and parameters. Specifically, we use a head setting to reserve essential information at the start of the network and propose a particular embedding setting to reduce parameters of fully-connected layer for embedding. To enhance interpretation ability, we employ an equivalence of angular distillation loss to guide our lightweight network and we apply recursive knowledge distillation to relieve the discrepancy between the teacher model and the student model [47].

Table 1 comparison of different face databases

Database	Description	Limitation
AT&T [12]	Contain face image of 40 persons, with 10 images of each. For most subjects the 10 images were shot at different times and with different lighting conditions, but always against a dark background.	1, limited number of people 2, illumination conditions are not consistent from image 3, the images are not annotated for difference facial expressions
Yale [13]	Contains frontal gray-scale face images of 15 people with 11 face images of each subject giving a total of 165 images. Lighting variation include left-light, center-light and right-light.	1, limited number of people 2, the precise positions of the light sources are not specified 3, environmental factors are not described
MIT [14]	Contains 16 subjects each subject sat on a couch and was photographed 27 times while varied head orientation. The lighting direction and the camera zoom were also varied during the sequence.	1, although this database contains images that were captured with a few difference scale variations, lighting variation and pose variation

Table 2 A summary of the research results depending on different methods

Methods	Reference	Percentage of correct calcification	Gap Description
Eigenface	[5]	95% correct classifications average over lighting, orientation, and size variation, respectively	The images are including a large quality of background area
SVM	[6]	Performance of identification is 77.78 % Performance of verification is 93 %	SVM was used only with polynomial (up to degree 3) and Gaussian kernel
SVM+3D morphable model	[7]	Up 98% verification	Face rotation up to 36 ⁰ in depth will rotate
SVM+PC+LD	[8]	Verification is 99% and recognition is 98%	It uses small amount of database

FaceNet	[9]	96.43%	The database contains frontal faces under controlled condition It takes long time for training It requires high standard of operation environment
SVM+PCA	[10]	92.67%	SVM was used only with polynomial up to degree 3 and Gaussian kernel

These papers show that there is a need for in-depth research on face recognition for more accurate detection and classification. Further, it has been observed while reviewing the literature of research on face image Segmentation that proposed methods work well only on face images that were captured in a controlled environment. Feature extraction is a mainly used technique used by researchers and there is a need for in-depth research on face recognition for more accurate detection and classification. The proposed methods and algorithm are used to solve the gaps of these finds.

The papers use different types of Facial recognition algorithms. they produce some errors like: false positives and false negatives. The first occur when the algorithm thinks there's a positive match between two facial images, but in fact there is no match. The second take place when the algorithm says there's no match, but in fact there should be one. And depending on different databases give different results of accuracy.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Overview

The purpose of this chapter is to introduce the methodology we used for our experiments that contributed to the final outcomes. This includes data sources data augmentation method and model design. We will clearly explain the details and process of the method.

Collect human faces by capturing and from the database. And preparing the necessary data set by using different mechanisms such as image preprocessing, extracting and augmentation. Based on the prepared dataset Convolutional Neural network (CNN) algorithm detect either face or not. The sources for the kinds of literature going to review will be included Springer Link, IEEE Explore, Science Direct, Academia, and Google scholar. The algorithm written below illustrated the step-by-step approach for the proposed Face recognition and detection.

After collecting individual faces and create a dataset. Design a convolutional neural network model to detect and classify the individual faces. In addition to that train the proposed model by using the collected face dataset and validate and test the proposed model with validation dataset. Validate and test the proposed model with a validation dataset.

3.2 Data collection

The data set by collected using a digital camera and using other online datasets. To collect the required data capturing the individual images will be undertaken. The data contains images with several Age variation images, Gender Female and male images, and different Ethnicity frontal face positions. For proposed research, all the required data will be collected for face recognition from different sources like images download from the internet or simply taking pictures using camera devices.

In deep learning data is the most major important thing that is used to develop the model. In deep learning huge number of datasets are needed for training. In this study two types of datasets are used for implementing the model which is the training set and testing set. Through these processes total collected data for the experiment is about 19108 images.

3.3 Development Tools

3.3.1 Software Tools

Anaconda IDE Navigator

Anaconda is an open source and free distribution of R and Python programming language for a machine learning as well as data science project.

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages environments and channels without using command line. Anaconda has more than 7,500 Anaconda built data science or machine learning packages.

Spyder the Scientific Python Development Environment is a free integrated development environment (IDE) that is included with Anaconda. It includes editing, interactive testing debugging, and introspection features.

Anaconda environment has the following graphical python application are already installed on the home tab or is available to install: JupyterLab, Jupyter notebook, Orange data visualization, Spyder IDE, Glueviz multidimensional data visualization, R studio IDE, PyCharm, VS Code, Anaconda Prompt, Anaconda PowerShell, and so on. Anaconda supported libraries or packages are Numby, Scipy, Scikit learn, Matplotlib, NLTK, Pandas, Bokeh, R essential.



Figure 2 Diagram of Anaconda Navigator

OpenCV library also used for Anaconda

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision application and accelerate the use of machine perception in the commercial products.

The OpenCV has more than 2500 optimized algorithms libraries which includes a comprehensive set of both classic and state of the art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces identify objects classify human action in videos, track camera moments, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high-resolution image of an entire scene, find similar image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 14 million. The library is used extensively in companies, research groups and by the governmental bodies [22].

It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision application and takes advantage of MMX and SSE instructions when available. A full featured CUDA and OpenCV interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a template interface that works seamlessly with STL container [22]. This why select for study.



Figure 2 Diagram of OpenCV

Face Detection and Recognition takes different approach and algorithm to detect and recognize human faces. There are different methods for face recognition which are the following

1. Geometric based/ Template based

The geometric feature-based methods analyse local facial features and their geometric relationship. It is also known as a feature-based method. Face recognition algorithm classified as geometry based or template-based algorithms. The template-based methods can be constructed using statistical tools like SVM, PCA and LDA.

2. Piecemeal/ Holistic

The relation between the elements or the connection of a function with the whole face not undergone into the amount many researchers followed this approach trying to deduce the most relevant characteristics.

3. Appearance Based / Model Based

The appearance-based method shows a face regarding several images. An image considered as a high dimensional vector. This usually used to drive a feature space from the imaged division.

4. Template / Statistical/ Neural Networks Based

4.1.Template Matching

In the template matching the patterns are represented by samples, models, pixels textures. The recognition function is usually a correlation or distance measure.

4.2.Statistical Approach

In the statistical approach is the patterns expressed as features. The recognition function in a discriminant function. The goal is to choose and apply the right statistical tool for extraction and analytical. Example of statistical approaches are the following: -

Principal Component Analysis (PCA)

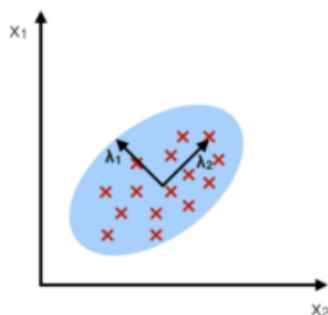
A mathematical procedure performs a dimensional reduction by extracting the principal components of multi-dimensional data.

Linear Discriminant Analysis (LDA)

LDA is widely used to find the linear combination of features while preserving class separability.

PCA:

component axes that maximize the variance



LDA:

maximizing the component axes for class-separation

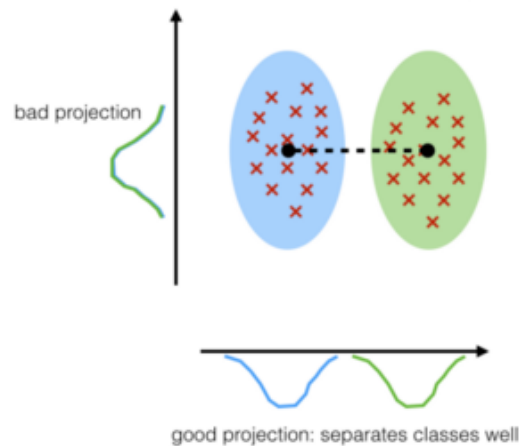


Figure 3 PCA vs LDA

Several references show the effectiveness of this system in terms of good detection.

1. **Integral Image:** The integral image is defined as the summation of the pixel values of the original image. The value at any location (x, y).
2. **Haar Features:** Haar features are composed of either two or three rectangles. Face candidates are scanned and searched for Haar features of the current stage. The weight

and size of each feature and the features themselves are generated by the learning algorithm. Each Haar feature has a value that is calculated by taking the area of each rectangle multiplying each by their respective weights and then summing the results.

3. **Classifier:** A Haar classifier uses the rectangle integral to calculate the value of a Haar feature. Depending on the parameters of the training data individual's stages can have a varying number of Harr features.
4. **Cascade:** The cascade eliminates candidates by making stricter requirements in each stage with later stage being much more difficult for a candidate to pass.

3.3.2 Hardware Tool

Digital camera was used to capture the sample individual images. To implement the Haar cascade algorithm with the selected software tools a machine with CPU Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz 2.30 GHz processor, RAM 4 GB.

3.4 Image Acquisition

Image acquisition is the process of obtaining a digital image from a different camera type. This capture used frame grabber which captures single frames of video and converts the analogue values to digital and feeds the result into the computer memory [25]. The captured images are stored in frame. Some of the image is stored in image frame and other is stored in test image. Images are used for training and other for test. And the machines or computer is learning every new image from the training and test.

3.5 Image Data Pre-processing

Our collected dataset is RGB or gray scale image. As most of the neural network models assume a square shape input image, they are uniform aspect pixel ration. Removing low-frequency background noise normalizing the intensity. Haar feature make image processing involved cropping of all the images and capturing from real time making the square around the faces and ensured that the image contains all the required information for feature extraction.

3.6 Image Data Augmentation

Data Augmentation is a technique that allows us to generate large data sets from small ones. We may not be able to obtain enough data for our study for a variety of reasons. Deep learning is a type of methodology that works effectively when we only have a large data collection. As a

result, we will need to supplement the small quantity of data about 19108 we originally acquired to build a good model. Adding effects such as enhancing contracting, flipping, and rotating at different angles some of the ways that can be used to accomplish this. This process of augmentation allows us to generate many alternative data samples from a single data set.

In most cases of image classification research, the dataset is very large. Data augmentation is a process of increasing the number of training data points in a dataset by generating more data from the RGB original dataset [29] it is important to increase the number of datasets. It helps the network to learn more complex features from the data and prevent the problem of overfitting. In this study various data augmentation technique has been performed on the original image's dataset.

In the image processing we also need to alter the position of the bounding box because we are using the elliptical bounding box. The focus of this problem is on how to calculate the corresponding parameters of the ellipse after the transformation. Then we will introduce the calculation of this ellipse transformation.

The ellipse of the major axis on the x -axis x^2

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 (a \geq b) \dots \dots \dots (3.1)$$

Where a is the semi major axis of the ellipse and b is the semi minor axis of the ellipse. When $a=b$ the ellipse becomes a circle. This elliptical parametric equation can be written as

$$\begin{cases} x = acost \\ y = bsint \end{cases} t \in [0, 2\pi] \dots \dots \dots (3.2)$$

Now we rotate the ellipse shown in equation (3.2) set the angle between the long axis of the ellipse and the x -axis as $\emptyset$ then the parameter equation can be written as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\emptyset & -\sin\emptyset \\ \sin\emptyset & \cos\emptyset \end{pmatrix} \begin{pmatrix} acost \\ bsint \end{pmatrix} \dots \dots \dots (3.3)$$

Then we shift center of the ellipse to (x_c, y_c) and finally get the parameter equation for the ellipse.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\emptyset & -\sin\emptyset \\ \sin\emptyset & \cos\emptyset \end{pmatrix} \begin{pmatrix} acost \\ bsint \end{pmatrix} + \begin{pmatrix} x_c \\ y_c \end{pmatrix} \dots \dots \dots (3.4)$$

3.7 Feature Extraction

Feature extraction is the main stage of image classification. The important features that are used to classify the image are extracted by the CNN algorithm. In the existing system using some features parameters like color feature, in this proposed model it automatically extracts features

from face recognition datasets. Deep learning enables machines to learn from experience and understand. Deep learning learns pathogens and classifiers naturally not like machine learning.

3.8 Performance Evaluation Metrics

After implementing the system evaluation is necessary to check the system is met the proposed objective or not is the next step. There is different measurement were used to assess the model's success rate and efficiency. Here are the following confusion metrics are used to evaluate implemented model.

Table 3 Confusion matrix

Truth\prediction	Actual positive	Actual negative
Prediction positive	True positive (TP)	False positive (FP)
Prediction negative	False Negative (FN)	True Negative (TN)

A confusion matrix is an evaluation statistic for describing a classification tasks performance. A confusion matrix can be used to calculate precision, Fallout, Recall and Specificity.

Accuracy shows the proportion of true result, both true positives and true negative.

$$\text{Precision} = \frac{TP}{TP+FP} \dots \dots \dots 3.1$$

$$\text{Fallout} = \frac{TN}{TP+TN} \dots \dots \dots 3.2$$

$$\text{Recall} = \frac{TP}{TP+FN} \dots \dots \dots 3.3$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \dots \dots \dots 3.4$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Recall} + \text{precision}} \dots \dots \dots 3.5$$

Where TP: True Positive are all samples that are classified as the correct class.

FP: False Positive are all samples that are classified as the wrong class.

TN: True Negative are all samples that are classified as unknown correctly.

FN: False Negative are all samples that are classified as unknown wrongly.

From equation (3.1- 3.5) the Face Recognizer compare the images of all individual with respect to all images in side of face database. The True Positive is then the accuracy is 98.6. this does not mean that it often accurate under all condition. It depends on the quality of camera and images in training and testing face. If the image face in both training and testing are taken with not high-quality camera and face database, the percent of False Positive (FP) increase and the degree of accuracy decrease.

CHAPTER FOUR: RESEARCH DESIGN

4.1. Introduction

This chapter focuses on the design of the proposed work and its experimental parameters. Specifically, the design of the proposed model and descriptions, how features are extracted, and classification is performed in the proposed model and other pertained models by using the technique called training from scratch are described briefly.

4.2 Model Selection

A deep learning algorithm which is CNN is chosen by using different literature that was conducted in computer vision especially in image classification. CNNs represent an interesting method for adaptive image processing. And also, the algorithm is used for feature of extraction, classification, training, testing as well as for evaluating the accuracy of the model. Convolutional neural network takes a data without the need for separate per-processing or feature extraction stage and with the preprocessing data. In addition to these feature extraction and classification stages occur naturally within a single framework.

4.3 Model Design

A deep learning algorithm which is CNN is chosen on different literature that was conducted in computer vision especially in image classification. CNNs represent an interesting method for adaptive image processing. The algorithm is used for feature extraction, classification, training, testing as well as for evaluating the accuracy of the model. CNN's take a data, without the need for separate pre-processing or feature extraction stage and with the preprocessing data. In addition to these, feature extraction and classification stages occur naturally within a single framework.

The key advantage of utilizing the CNN algorithm used in the TomNet model for detecting and classifying Face Recognition and Detection is that it is increasingly automated than traditional machine learning algorithms. The necessity to design multiple algorithms for different issues in the conventional machine learning algorithm requires the usage of more handcrafted algorithms, while on CNN of this model, we just need to design one algorithm for all issues. TomNet is the convolutional neural network model named by myself. This model is created from the scratch

using a total of 19108 augmented data individual images which are categorized into four classes. These classes are containing **Age, Sex, and Ethnicity** which are common types of face recognition. The total collected dataset is divided into five training set, validation, and test data set. Different division methods are used to split the dataset into subsets. For example, 70%, 15%, and 15 %, for training, validation, and test, respectively.

This network structure is mainly divided into the following steps:

- First, we have to input the face images
- Secondly the input face images are extracted through the CNN structure for feature extraction
- Finally, for CNN input the model uses two structures to achieve object classification and localization

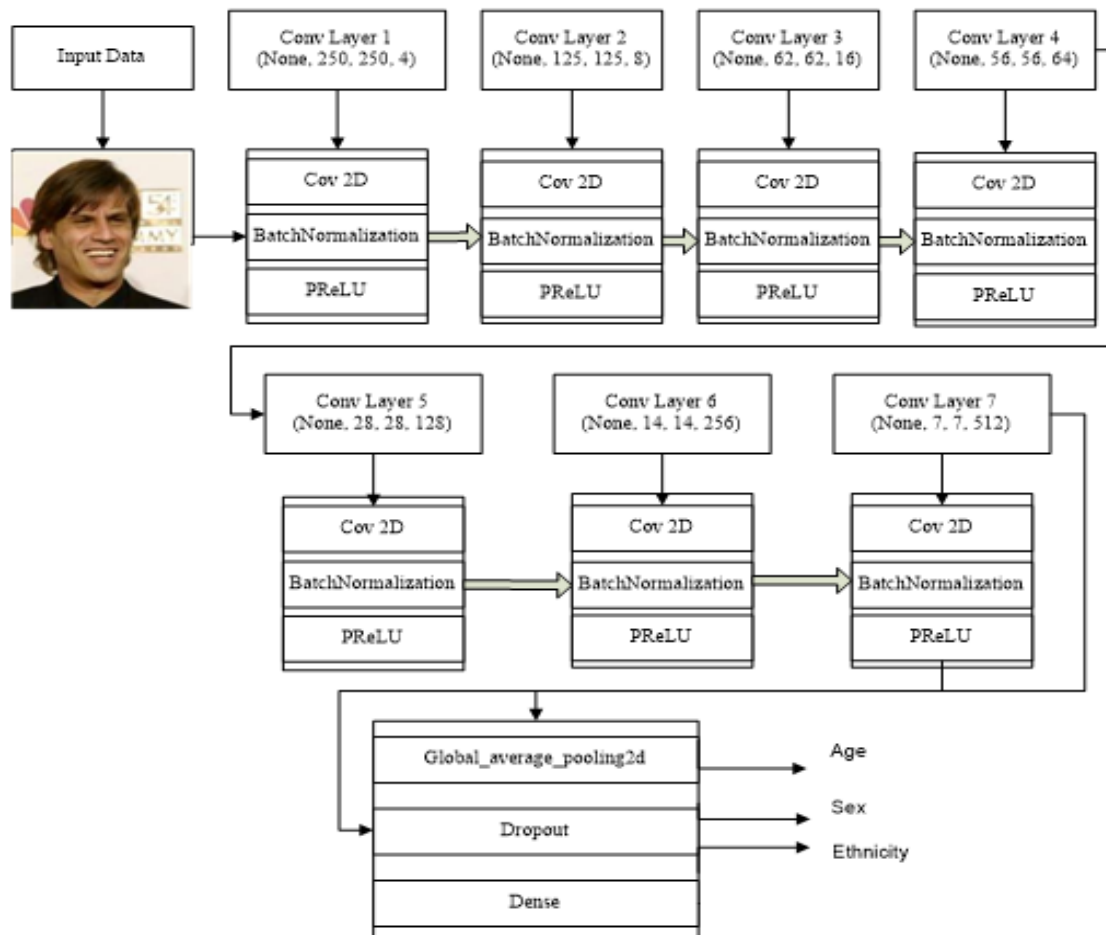


Figure 4 Face Recognition and Detection Diagram

4.3.1 Input Layer

The input layer of TomNet model accepts gray scale and RGB images of size 250*250*3 with three different classes (1, Age, 2, Gender, 3, Ethnicity) this layer only passes the input to the first convolution layer without any computation. Therefore, there are no learnable features and the number of parameters in this layer is 0.

4.3.2 Convolutional Layer (Convlayer)

The convolutional layer is the main and first layer of the CNN architecture. It is used to extract the characters from the input image. The convolution layer maintains the association connects pixels by learning image features from input data. The mathematical operation takes two input image matrices and 3x3 kernel filters. The 5x5 image pixel values and the convolution filter of 3x3. The CNN adds and divides the result by the total number of pixel and then create a map and put the value of the filter at that place. After that, it moves the feature to every other position of the image and obtains the output of the matrix, and repeats the steps for the other filters. This layer moves the filter to every possible position on the image.

4.3.3 Pooling Layer

This layer compresses the image into a smaller size by taking the maximum value from the filtered image and obtains a matrix out of it. It also controls over fitting. In the TomNet model, there are three max-pooling layers.

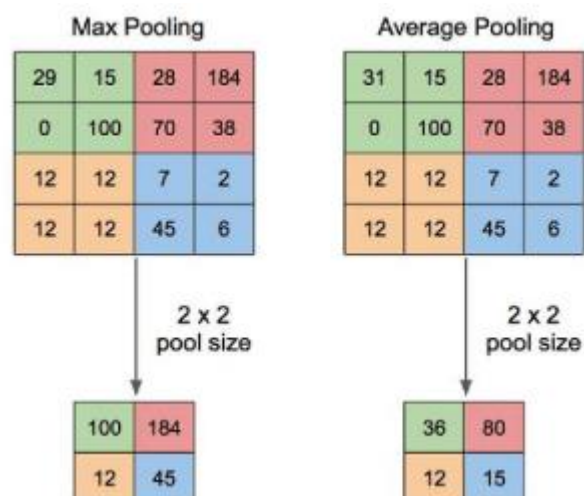


Figure 5 Max Pooling and Average Pooling

4.3.4 Fully Connected (FC) Layer

In the TomNet model, there are two fully connected layers including the output layer. The first fully connected layers have 64 neurons each and the final layer which is the output layer of the model has only one neuron. The first FC layer accepts the output of the seven Convolution layer after converting the 3D volume of data into a vector value (Flattening). This layer computes the class score and the number of neurons in the layer predefined during the development of the model. It is the same as ordinary neural networks and as the name implies, each neuron in this layer is connected to all the numbers of layers.

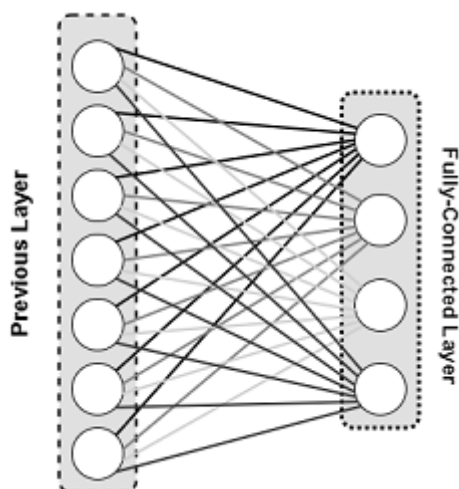


Figure 6 Full Connected layer

4.3.5 Output Layer

The output layer is the last (the third FC layer) of the model and it has 3 neurons with a SoftMax activation function. Because the model is designed to classify four classes called categorical classification.

Table 4 TomNet model summary

Model: "Sequential"		
Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, 250, 250, 4)	112

norm_1 (BatchNormalization) (None, 250, 250, 4)	16
max_pooling2d (MaxPooling2D (None, 125, 125, 4)	0
)	
conv2d_1 (Conv2D) (None, 125, 125, 8)	296
norm_2 (BatchNormalization) (None, 125, 125, 8)	32
max_pooling2d_1 (MaxPooling (None, 62, 62, 8)	0
2D)	
conv2d_2 (Conv2D) (None, 62, 62, 16)	1168
norm_3 (BatchNormalization) (None, 62, 62, 16)	64
max_pooling2d_2 (MaxPooling (None, 31, 31, 16)	0
2D)	
conv2d_3 (Conv2D) (None, 31, 31, 32)	4640
norm_4 (BatchNormalization) (None, 31, 31, 32)	128
max_pooling2d_3 (MaxPooling (None, 15, 15, 32)	0
2D)	
conv2d_4 (Conv2D) (None, 15, 15, 32)	25632
norm_5 (BatchNormalization) (None, 15, 15, 32)	128
max_pooling2d_4 (MaxPooling (None, 7, 7, 32)	0
2D)	
conv2d_5 (Conv2D) (None, 7, 7, 64)	18496

norm_6 (BatchNormalization)	(None, 7, 7, 64)	256
max_pooling2d_5 (MaxPooling	(None, 3, 3, 64)	0
2D)		
conv2d_6 (Conv2D)	(None, 3, 3, 64)	102464
norm_7 (BatchNormalization)	(None, 3, 3, 64)	256
max_pooling2d_6 (MaxPooling	(None, 1, 1, 64)	0
2D)		
conv2d_7 (Conv2D)	(None, 1, 1, 150)	86550
norm_8 (BatchNormalization)	(None, 1, 1, 150)	600
max_pooling2d_7 (MaxPooling	(None, 1, 1, 150)	0
2D)		
conv2d_8 (Conv2D)	(None, 1, 1, 150)	22650
norm_9 (BatchNormalization)	(None, 1, 1, 150)	600
max_pooling2d_8 (MaxPooling	(None, 1, 1, 150)	0
2D)		
global_average_pooling2d (G	(None, 150)	0
lobalAveragePooling2D)		
activation (Activation)	(None, 150)	0
dropout (Dropout)	(None, 150)	0
dense (Dense)	(None, 5)	755

activation_1 (Activation) (None, 5)	0
=====	
Total params: 264,843	
Trainable params: 263,803	
Non-trainable params: 1,040	

4.4 Dataset

The data has been collected from a different place as discussed in a methodology. The dataset used in this study has four classes. These are the common types of face recognition. Which are namely Age, Sex, and Ethnicity.

There are a total of 3 kinds of features that are used to categorize input images the data structure format is shown as follows:

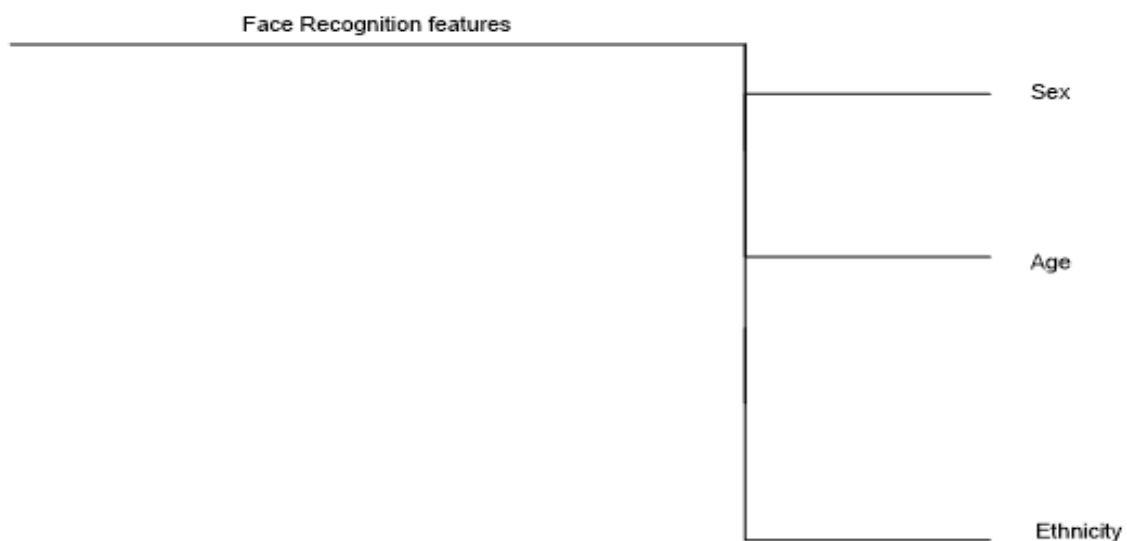


Figure 7 Input structure of Dataset

4.5 Sample Input Image Dataset

The images are collected in a dataset from different sources. The all types of images are categorized into one folder. Then when the model is training if any new input data comes having the same features as of these data images, then that image will also get included in the existing dataset. And the model gets smarter by the new features of the input image data is shown in figure 3.4.

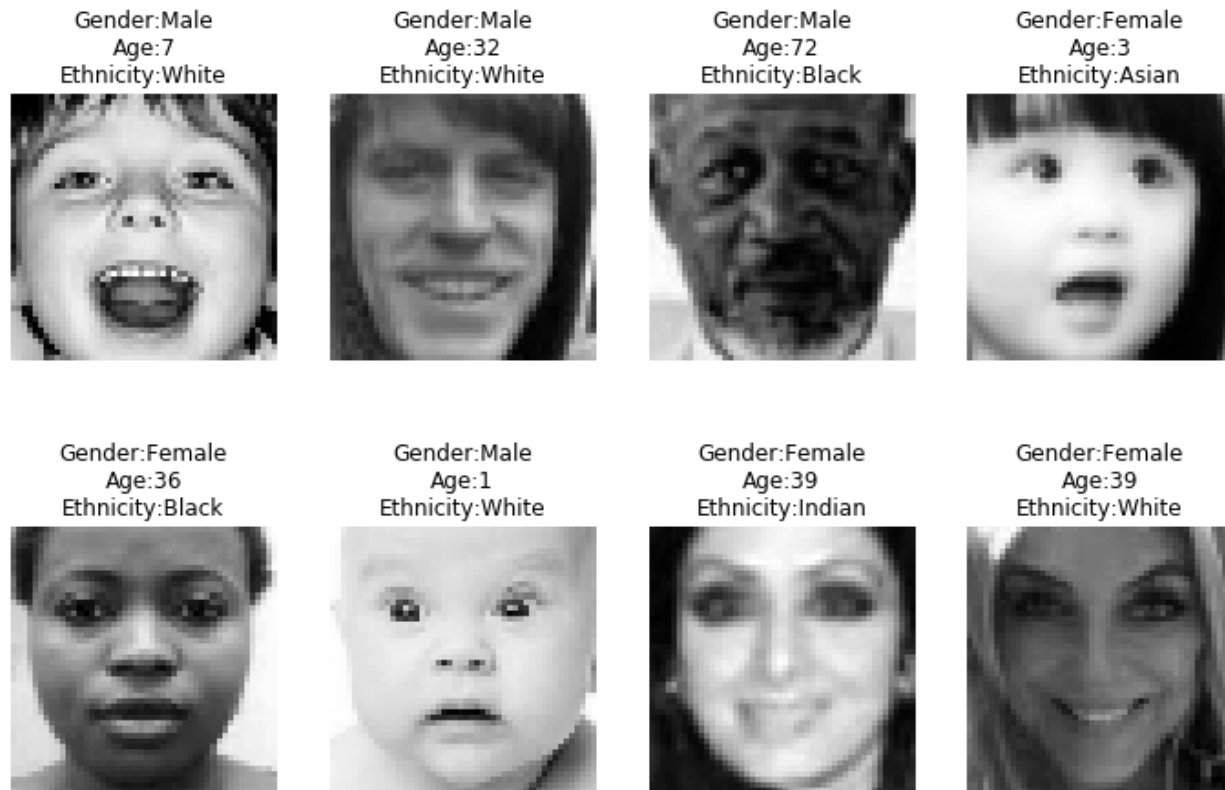


Figure 8 Example of the training image Database

4.6 Image conversion to gray-scale

A digital image is an image composed of picture element also known as pixels each with finite discrete quantities of numeric representation for its intensity or gray level that is an output from its two-dimensional function fed as input by its spatial coordinates donated with x, y on the X-axis and Y- axis. If you remove the color information you are left with gray-scale resulting in a black and white image. Gray-scale is a range of shades of gray without apparent color. The darkest possible shade is black which is the total absence of transmitted or reflected light. Gary is

represented by equal brightness levels of the three primary colors (Red, Green, Blue) for transmitted light.

A digital image is composed of groups of three pixels with colors of Red, Green, Blue (RGB) also called in digital imaging. RGB images are usually used in computer monitors and scanners.

This is OpenCv python code to take image from Real- Time

```
while True:

    # Capture frame-by-frame

    ret, frames = video_capture.read()

    gray = cv2.cvtColor(frames, cv2.COLOR_BGR2GRAY)
```

why we use gray scale image is

A grayscale image is simply the only color that is shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel. In addition

1. Signal to noise. for many applications of image processing color information doesn't help us identify important edges or other features. If there is an edge in hue that is hard to detect in a gray-scale image or if we need to identify objects of known hue, then color information could be useful. If we don't need color, then we can consider it noise.
2. For learning image processing it's better to understand gray-scale processing first and understand how it applies to multichannel processing rather than starting with full color imaging and missing all the important insights that can be learning from single channel processing.
3. Complexity of the code. If you want to find edges based on luminance and chrominance you've got more work ahead of you. That additional work and debugging, pain in supporting the software is hard to justify if the additional color information isn't helpful for applications of interest.
4. Difficulty of visualization. In gray-scale images the watershed algorithm is fairly easy to conceptualize because we can think of the two spatial dimensions and one brightness dimension as a 3D image with hills valleys catchment basins, ridges etc. one workaround

is to think of each color variable as an intensity image but that leads us right-back to gray-scale image processing.

5. Color is complex. Human perceive color and identify color with deceptive ease.
6. Speed. With modern computers and with parallel programming it's possible to perform simple pixel-by pixel processing of a megapixel image in milliseconds. Facial Recognition OCR content aware resizing mean shift segmentation and other task can take much longer than that.

4.7 Algorithms for Optimization

TomNet model is trained using a gradient descent optimization algorithm to minimize the error rate and the back-propagation algorithm is used to update the weights. Gradient descent is the most popular and the most widely used optimization algorithm in deep learning researches [23]. At the same time, every state-of-the-art Deep Learning library contains implementations of gradient descent optimization algorithms such as Keras and Tensorflow (used in this thesis). It updates the weight of the model and tunes parameters, therefore, minimize the loss function. The Adaptive Moment Estimation (Adam) optimizer is used to optimize the gradient descent. The most common optimization algorithm used today for training deep neural networks is Adam to computes adaptive learning rates. As it is straightforward to implement, computationally very efficient, and is very effective in dealing with large data and parameters. Each parameter and uses squared gradients to scale the learning rate; it also uses the moving average of the gradient.

4.8 Activation function

Experiments are conducted by using two different activation functions: Softmax and Sigmoid are used the TomNet model. Especially, in the output layer of the model, the sigmoid activation functions.

4.8.1 Epochs

It is the number of times the whole training data is shown to the network while training. It is the number of iterations the entire training dataset passes forwarded and back warded through the model or the network. This model was trained by using epochs starting from 10. This was done to get the best optimal epoch.

4.8.2 Batch size

It is the total number of training input passed into training. The single set is divided into small subsets called mini-batches to update the parameters. Though, to prevent any loss of accuracy and keep the generality of the model, the model is trained in the whole dataset.

4.8.3 Learning rate

Back-propagation was used to train the TomNet model; the learning rate is used during weight updates. It controls the amount of weight to update during back-propagation. The challenging part was to choose the proper learning rate during the experiment. The learning rates with a value of too small take longer to train than a value of larger. But when given a smaller value the model is more optimal than a model with a learning rate of larger. The experiment was done by using a learning rate of 0.01, 0.001, and 0.1. This was done to pick the best optimal learning rate.

Table 5 Summary of hyper-parameters used during model training

<i>Parameter</i>	<i>Value</i>
Activation Function	Softmax, Sigmoid
Optimization algorithm	Adam, SGD
Loss Function	BCE, CCE
Epoch	10
Bach size	32-64
Learning rate	0.001,0.01 and 0.1

4.9 Hyper-parameters for TomNet Model

All the convolutional neural network models employed using various parameters. These parameters make so many possible changes in the architecture. Training them with the huge dataset needed quite a long time. Performing hyper parameter optimization is used to makes efficient computing with resources that available. Thus, we utilized some common hyper parameter values and looked further into techniques to achieve a better model evaluation. The following are hyper-parameters that are chosen for the TomNet model;

- **Dataset types**

RGB images

Gray scale images

RGB Augmented images

- **Deep learning architecture:**

TomNet

- **Training mechanism**

Training from Scratch

- **Training testing set distribution**

Train: 80%, Test: 20%,

Train: 70%, Test: 15%, validation 15%

Train: 70% Test: 30%

4.10 Histogram Equalizer

Histogram equalizer is used to enhance contrast. It is not necessary that contrast will always be increase in this. There may be some cases where histogram equalization can be worse. In those cases, the contrast is decreased [46]. Histogram is the estimation of the probability distribution of a particular type of data. An image histogram is a type of histogram which offers a graphical represent.

CHAPTER FIVE: IMPLEMENTATION

5.1 Proposed System

In this topic describes how the entire implementation was done. And all the process passed to get the research result is described in detail. It contains data preparation, data augmentation, data split and image processing and building the model.

5.2 Data Preparation

Many researches are done on Face Detection and Recognition on different sectors. From these sectors one is Bank. The Banking activity needs more security so they change the system to biometric system. Different Face Detection and Recognition algorithm determine the quality and effectivity as well as the accuracy of Detection and Recognition. The researchers' works on automated face recognition for different machines have ignored the issue of just what aspects of the face stimulus are important for face recognition.

The images have different proportions; their size was modified to a 250x250 pixel proportion. This resolution was the same satisfactory for both maintaining the image features, as well as keeping computational time to a minimum. All individual images were originally saved in an RGB 3-channel. To achieve the efficient model as required the first thing that must be done is sort out the data set. Before train the model it is important to well prepare a proper sufficient data set. Originally the collected individual face images from the database are not enough to use for deep learning. So, the first thing done to increase the data set size is using a Data Augmentation. This is used to enhance the size of the original data. For this study, the dataset is a total of 19108.

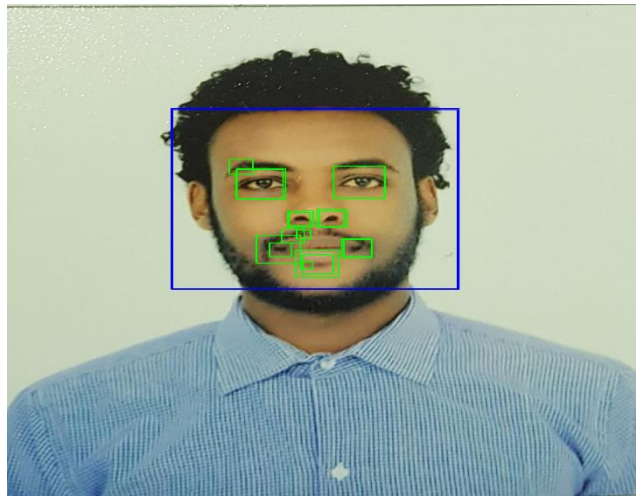


Figure 9 Sample original image

These original data images are multiplied using different features though augmentation processes sample of augmented images are the following.

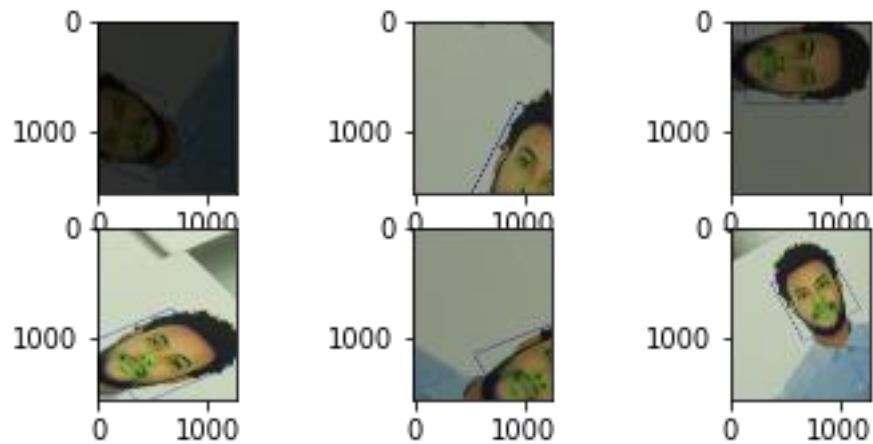


Figure 10 Augmented Images

Sample Code for Data Augmentation

```
from numpy import expand_dims
from keras.preprocessing.image import load_img
from keras.preprocessing.image import img_to_array
from keras.preprocessing.image import ImageDataGenerator
from matplotlib import pyplot
# Load the image
img = load_img('tomi.png')
# convert to numpy array
data = img_to_array(img)
# expand dimension to one sample
samples = expand_dims(data, 0)
#image data augmentation generator
datagen = ImageDataGenerator(width_shift_range=[-200,200],
                             zoom_range=[0.5,1.0],
                             brightness_range=[0.2,1.0],
                             rotation_range=90,
                             horizontal_flip=True,
                             height_shift_range=0.5,
                             fill_mode='nearest')
```

Figure 11 Sample code for data augmentation

5.3 Image Preprocessing

The images inside of the training data have a sparse nature with different dimension sizes, thus it needs to normalize to the same image dimension, which is making the training data standardized to the same image dimension size. I have used a dimension of 250*250 for all images in the training set. Images are loaded recursively based on their pathname. Depending on the prefix of their filename, a label value of '0' or '1' is appended to each loaded image. Color conversion: Two functions for color conversion were implemented in this thesis: a conversion from the BGR to RGB color-space, where the order of the image's color-space, where the images lose all the color information and maintain only the brightness and saturation for each pixel.

Data normalization: is achieved by finding the difference of the mean from every pixel, after that separating the outcome by the standard deviation. The dataset values are normalized within the (0, 1) range to ensure that the loss function, which is usually not convex, finds the global minimum as easily as possible. Reducing the range of input for training values also assists the convergence of the back-propagation algorithm.

Dataset shuffling: The dataset is being shuffled by python's randomly Shuffle method, which is an order-shuffling algorithm based on a random number generator. After the application, the order of the images is originally being alphabetical, now mixed throughout the dataset.

Dataset splitting: the data were split into different fashion such as the first 15% of the total of the dataset is held out and set as a testing dataset, the second 15% of data has set to the validation, and the remaining dataset is splitting again 80/20 fashion, where the 20% portion is used as the validation dataset. The rest is used for training dataset the algorithm. The training and validation sets are used for the training of the algorithm, while the testing dataset is used only after the classifier is trained, to make predictions. The following are sample preprocessed images.

Model: "sequential_3"		
Layer (type)	Output Shape	Param #
conv2d_21 (Conv2D)	(None, 258, 258, 128)	3584
norm_1 (BatchNormalization)	(None, 258, 258, 128)	512
max_pooling2d_21 (MaxPooling2D)	(None, 125, 125, 128)	0
conv2d_22 (Conv2D)	(None, 125, 125, 256)	295168
norm_2 (BatchNormalization)	(None, 125, 125, 256)	1024
max_pooling2d_22 (MaxPooling2D)	(None, 62, 62, 256)	0
conv2d_23 (Conv2D)	(None, 62, 62, 16)	36880
norm_3 (BatchNormalization)	(None, 62, 62, 16)	64
max_pooling2d_23 (MaxPooling2D)	(None, 31, 31, 16)	0
conv2d_24 (Conv2D)	(None, 31, 31, 32)	4640
norm_4 (BatchNormalization)	(None, 31, 31, 32)	128
max_pooling2d_24 (MaxPooling2D)	(None, 15, 15, 32)	0
conv2d_25 (Conv2D)	(None, 15, 15, 32)	25632
norm_5 (BatchNormalization)	(None, 15, 15, 32)	128
max_pooling2d_25 (MaxPooling2D)	(None, 7, 7, 32)	0
conv2d_26 (Conv2D)	(None, 7, 7, 64)	18496
norm_6 (BatchNormalization)	(None, 7, 7, 64)	256
max_pooling2d_26 (MaxPooling2D)	(None, 3, 3, 64)	0
conv2d_27 (Conv2D)	(None, 3, 3, 64)	102464
norm_7 (BatchNormalization)	(None, 3, 3, 64)	256
max_pooling2d_27 (MaxPooling2D)	(None, 1, 1, 64)	0
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 64)	0
activation_6 (Activation)	(None, 64)	0
dropout_3 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 5)	325
activation_7 (Activation)	(None, 5)	0
Total params: 489,557		
Trainable params: 488,373		
Non-trainable params: 1,184		

Figure 12 Sample code for TomNet model

Table 6 Sample Code for TomNet_model

Model: "sequential"		
Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, 250, 250, 128)	3584
norm_1 (BatchNormalization)	(None, 250, 250, 128)	512
max_pooling2d (MaxPooling2D)	(None, 125, 125, 128)	0
conv2d_1 (Conv2D)	(None, 125, 125, 256)	295168
norm_2 (BatchNormalization)	(None, 125, 125, 256)	1024
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 256)	0
conv2d_2 (Conv2D)	(None, 62, 62, 256)	590080
norm_3 (BatchNormalization)	(None, 62, 62, 256)	1024
max_pooling2d_2 (MaxPooling2D)	(None, 31, 31, 256)	0
conv2d_3 (Conv2D)	(None, 31, 31, 16)	36880
norm_4 (BatchNormalization)	(None, 31, 31, 16)	64
max_pooling2d_3 (MaxPooling2D)	(None, 15, 15, 16)	0
conv2d_4 (Conv2D)	(None, 15, 15, 32)	12832
norm_5 (BatchNormalization)	(None, 15, 15, 32)	128
max_pooling2d_4 (MaxPooling2D)	(None, 7, 7, 32)	0

conv2d_5 (Conv2D)	(None, 7, 7, 64)	18496
norm_6 (BatchNormalization)	(None, 7, 7, 64)	256
max_pooling2d_5 (MaxPooling2)	(None, 3, 3, 64)	0
conv2d_6 (Conv2D)	(None, 3, 3, 64)	102464
norm_7 (BatchNormalization)	(None, 3, 3, 64)	256
max_pooling2d_6 (MaxPooling2)	(None, 1, 1, 64)	0
global_average_pooling2d (Gl	(None, 64)	0
activation (Activation)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 5)	325
activation_1 (Activation)	(None, 5)	0
=====		
Total params: 1,063,093		
Trainable params: 1,061,461		
Non-trainable params: 1,632		

CNN is a class of deep neural networks commonly applied to analyzing visual imagery. Here is the TomNet models example of CNN with few layers using Conv2D -2D convolutional layer, Batch Normalization, and MaxPooling2D application of a moving window across 2D input space. Hyper parameters are set before training they represent the variables that determine the neural network structure and how it is trained. Conv2D with hyper parameters mentioned above: Conv2D (kernel size, (filters, and filters), input shape = (250, 250, and 3) with activation function for each layer as a ReLU.

- MaxPooling2D layer to reduce the spatial size of the incoming features;
- 2D input space: MaxPooling2D ((2, 2)) and ((2, 2))
- Do the same increasing the kernel size: 16-32-64-250
- Provide the last fully connected layer which specifies the number of classes of face leaf recognition.
- Print the summary of the model.

5.4 Haar features

Haar like feature are digital image features used for object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real time face detector. Historically working with only image intensities, the RGB pixel values at each and every pixel of image made the task of feature calculation computationally expensive. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. For example, let us say we have an image database with human faces. It is a common observation that among all faces the region of the eyes is darker than the region of the cheeks. Therefore, a common Haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek region. The position of these rectangles is defined relative to a detection window that acts like a bounding box to the target object.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature).

5.5 Description of Module

A. Capturing an Image

The proposed research is used either cameras or webcam of computer to capture face from real time object. Webcam is the normally the laptop frontal camera. The research take image or through two ways.

1. By capturing from real time video and
2. From external face database

B. Face Detection

In the face Detection and Recognition for ATM machine face detection is used to identify either it is face or not. This research is used Viola-Jones algorithm with Haar Cascade as Face detection a method. Viola-Jones detector was chosen as a detection algorithm because of its high detection rate, and its ability to run in real time [45]. Haar features can be various height and width. It makes some rectangular image on detected face. Haar feature can be of various height and width. This research is used 25×25 pixel. From the Haar feature applied to the face the sum of black pixel and sum of white pixel are calculated and they are subtracted to get a single value.

A Haar like feature however is far from enough to build a powerful nonlinear classifier. Cascade of classifiers makes the computation even more efficient which follows the principles of pattern rejection. Pattern rejection make program to take only face image [39].

Haar features are calculated all over the image which will be almost 160000+ features per image. Summing up the entire image pixel and then subtracting them to get a single value is not efficient in real time applications. This can be reduced by using Ada boost classifier. Ada boost reduces the redundant features. The integral image is an algorithm for cost effective generation of the sum of pixel intensities in a specified rectangle in an image of face detection. It is used for rapid computation of Haar like features. Calculation of the sum of a rectangular area inside the original image is extremely efficient requiring only four additions for any arbitrary rectangle size.

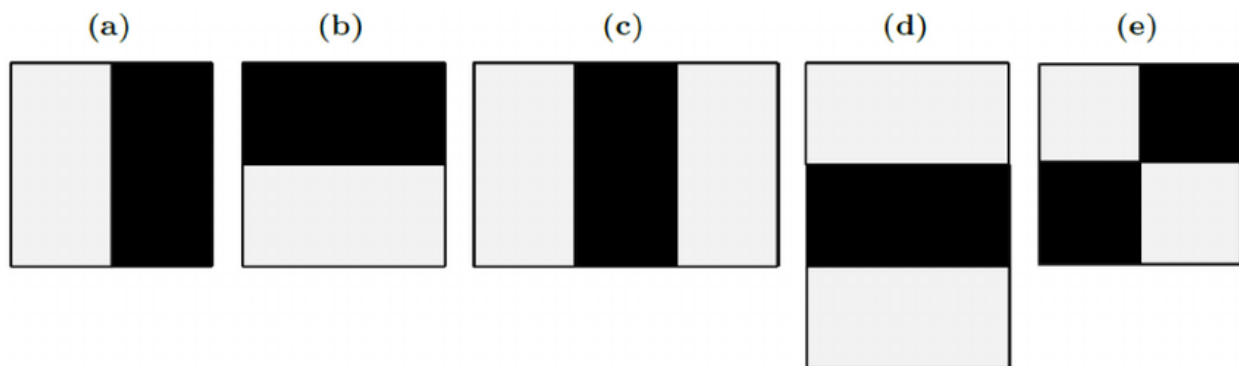


Figure 13 Haar Like Features in Viola Jones

5.5.1 Correct Face Position and Orientation

The study did not contact any form of correct position of an individual face either back or front as well as in any degree of inclination. Instead, person who are going to register in face database

are orient the customers position which can detected by camera. The face is detected only when the image is at correct position or Haar cascade frontal face.

5.5.2 Face Recognition

This stage is where bio-metric human individual face is recognized. The individual identity in the bank face database is identified either it is face database or not. If the face is not existing in the database there should re-post that face is not existing in the bank database. Otherwise, the real face is identified. In this study the face recognition is taken after training and testing of customers from the bank database.

In face recognition the data taken from the images are simulated using a previously trained method. The input will be a vector array from the previous stage. The networks are trained with face descriptors as input [36]. The number of networks will be equal to the number of persons in the database.

In a similar way the computer created artificial network has inputs for inserting the data outputs for providing the network output and hidden layer for processing the data and training of the network as shown in Figure4.5.

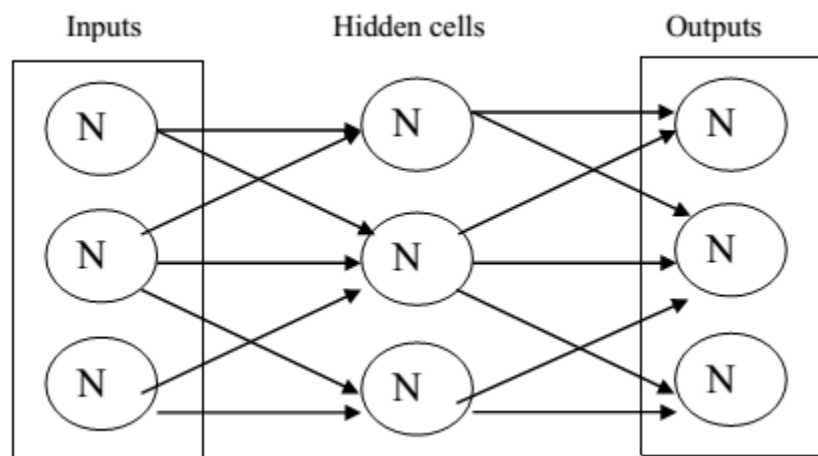


Figure 14 Artificial Neural Networks

5.5.3 Histogram Matching

Histogram equalization is a technique for adjusting image intensities to enhance contrast and determine the transformation that seeks to produce output images that has uniform histogram.

Histogram automatically determines the transformation that seeks to produce output images that has uniform histogram [30].

Let f be a given image represented as a m_r by m_c matrix of integer pixel intensities ranging from 0 to $L - 1$. L is the number of possible intensity values often 256. Let p denote the normalized histogram of f with a bin for each possible intensity:

$$P_n = \frac{\text{number of pixels with intensity } n}{\text{total number of pixels}} \quad n = 0, 1, \dots, L - 1 \dots \dots \dots (4.4)$$

The histogram equalized image g will be defined by:

$$g_{i,j} = \text{floor} \left((L - 1) \sum_{n=0}^{f_{i,j}} p_n \right) \dots \dots \dots (4.5)$$

CHAPTER SIX: RESULT, DISCUSSION AND EVALUATION

6.1 Discussion

Face Detection and Recognition in have limitation such as variations in pose, lighting, hair style, makeup and image quality. In addition, there is the need for a database to store images in order to recognize people. Obtaining good images of faces is not always easy. People could be wearing hats covering their faces, or be looking away from the camera. [30]

6.2 Result

After performing all the evaluation on the gray scale and RGB segmented dataset on the TomNet model experiments with top results are discussed below. The TomNet model is evaluated using different parameters that can affect the efficiency of the model. These parameters are several epochs; train test split ratio, learning rate, and dropout ratio which are explained concisely in previous chapters. Different parameter ratios are used for the evaluation purpose and the most successful ratios are discussed in this chapter.

In this thesis, three questions stated in the problem statement part are tried to be answered. To select the best model to solve the stated problem, the various experiments is conducted using CNN method. Using this method, the TomNet model is build using 38,076 (thirty-eight thousand seventy-six) training samples from 19,108 (Ninety thousand one hundred eight) total data set. This is 70% of the total dataset. Model fit is implemented using 18,968(eighteen thousand nine hundred sixty-eight) validation (testing) dataset from 9806 (Nine thousand eight hundred six) total dataset. This is 15% of the total dataset. In another experiment, the data was divided to 80% for training samples from the total dataset, 20% for test and validation data samples.

To come up with the best model the third try of the experiment was dividing the dataset into 60% for training samples and 40 % testing and validation data samples. The lack of a globally acknowledged standard ratio for separating the training and test data sets is the reason behind this. The TomNet model was built using RGB image training datasets and gray scale image datasets. Performances of both models are evaluated against validation datasets. The different scenarios are used in an experiment in addition to dividing datasets into training and testing sub-samples. The result obtained from the experiment done using RGB images discussed as follows in different perspectives. The experiment was done in different data splits as mentioned above.

These are 60-40, 70-30, and 80-20 the experiments are done accordingly. Using this different data partition, the result obtained is also different as listed in table.

Table 7 Experiment result using learning rate

No	Training samples	Test and Validation samples	Loss	Accuracy	Epochs
1	60%	40%	0.46	98.04%	10
2	70%	30%	0.012	98.62%	10
3	80%	20%	0.0014	98.88%	10

6.3 Result of Experiment done on RGB data images

Multi-channel images are displayed using the colors in RGB images (24-bit with 8-bits for each red, green, and blue channel). Colors have been chosen to reflect natural hues (i.e., the green in an RGB image reflects the green color in the specimen). As stated previously, this thesis is based on RGB and grey scale data images. The experiment is carried out utilizing various circumstances to improve the TomNet model's performance. The results of the experiments are presented in the tables below.

6.3.1 Experiment result using output Activation Function

The Sigmoid Activation Function is a mathematical function that has a distinct "S" curve. It is mostly used for logistic regression and the creation of simple neural networks. This function is used separately for each element's raw output in this thesis experiment. softmax is a technique for mapping non-normalized data output to a probability distribution for output classes. In neural network-based classifiers, it is used in the final layers. They are trained under either the log-loss or cross-entropy regime. Softmax often used in artificial and convolutional neural networks. These activation functions are used in the TomNet model. The following table 5 shows the results of the model experiment using these functions:

6.3.2 Experiment result using various Learning Rates

The learning rate is a hyper-parameter that governs how much the weights of our network are adjusted about the loss gradient. A value too little may result in a long training process that

becomes stuck, whereas a value too large may result in learning a sub-optimal set of weights too quickly or an unstable training process. The stochastic gradient descent algorithm estimates the error gradient for the current state of the model using examples from the training dataset, then updates the weights of the model using the back-propagation of errors process, sometimes known as simply back-propagation. The step size, often known as the "learning rate," is the amount by which the weights are changed during training. The learning rate is an adjustable hyper parameter that has a modest positive value, usually between 0.0 and 1.0, and is used in the training of neural networks. This experiment was done by using a learning rate of 0.1, 0.01, and 0.001.

Table 8 Experiment of the TomNet model using different learning rate

Learning Rate	Accuracy		Loss	
	Training	Validation	Training	Validation
0.1	94.45%	90.54%	20%	20.6%
0.01	94.32	91.43%	17.7%	13.4%
0.001	98.53	91.92%	4.3%	8.3%

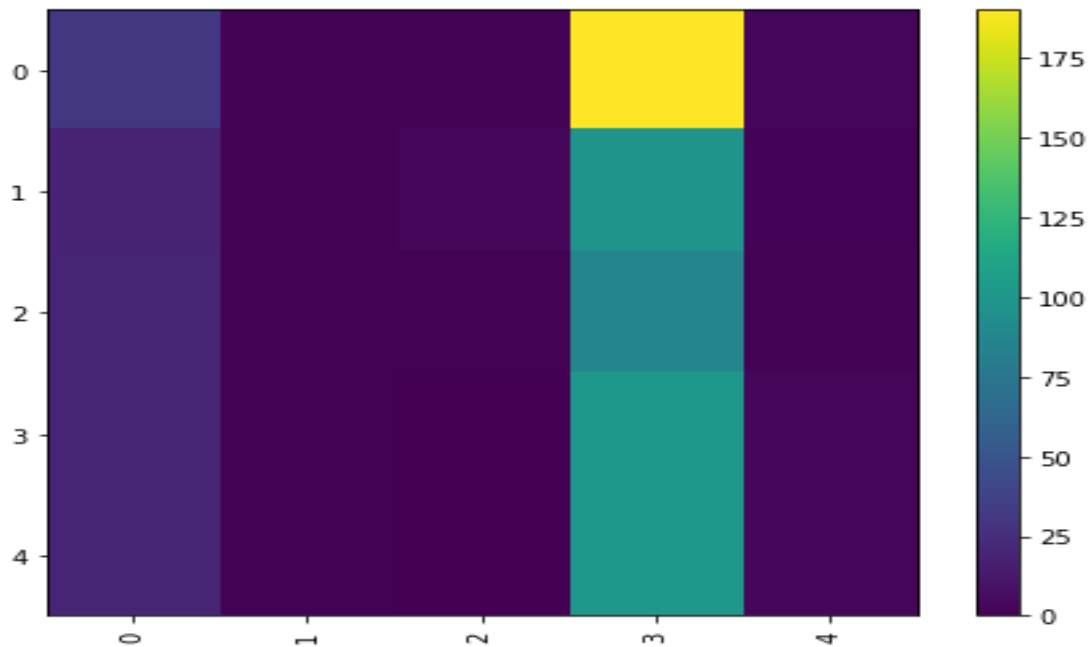
6.3.2.1 Activation Function

The result that is obtained from the experiment of TomNet model is by using different activation function is presented in the following tables by using the classification accuracy metrics in the form of percentage for the train, validation and test data separately.

table 9 Experiment result using different activation function on images

Activation function	Accuracy		Loss	
	Training	Validation	Training	Validation
Sigmoid	98.6%	98.83%	13.3%	12.3%
Softmax	93.8%	94.4%	0.005%	0.0074%

On the above table we can see the sigmoid activation is better than softmax by their layer of binary classification which is more preferred for multiclass classification problems. And the TomNet model was successfully classifies the given image with training accuracy of 98.6% and means Validation accuracy of 98.83% by using the learning rate of 0.001 output layer activation function of sigmoid and train and validation dataset ratio of 8:2.



Confusion Matrix					
	0	1	2	3	4
0	31	1	1	190	3
1	18	1	3	98	2
2	20	1	1	87	1
3	20	1	0	101	3
4	20	1	0	101	3
Classification Report					
	precision	recall	f1-score	support	
Age	0.28	0.14	0.19	226	
Sex	0.20	0.01	0.02	122	
Ethnicity	0.20	0.01	0.02	110	
Pose	0.18	0.81	0.29	125	
Smile	0.25	0.02	0.04	125	
accuracy			0.19	708	
macro avg	0.22	0.20	0.11	708	
weighted avg	0.23	0.19	0.12	708	

Figure 15 Confusion Matrix for CNN of images

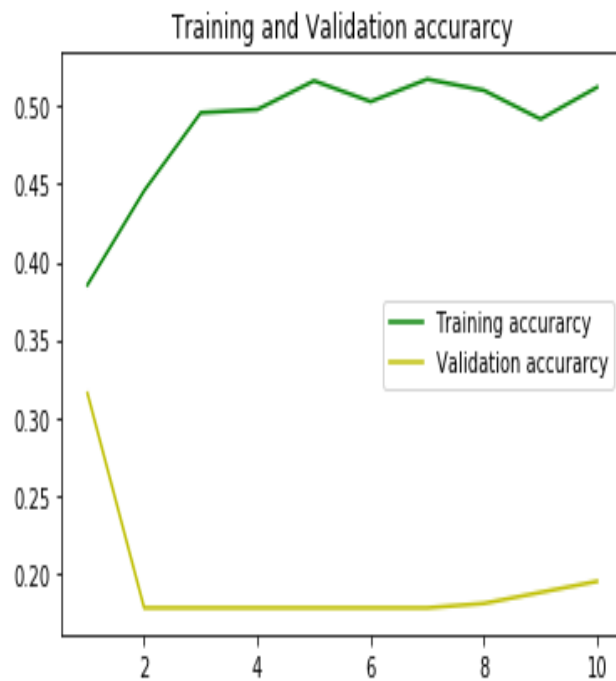


Figure 16 Training and Validation Accuracy Analysis

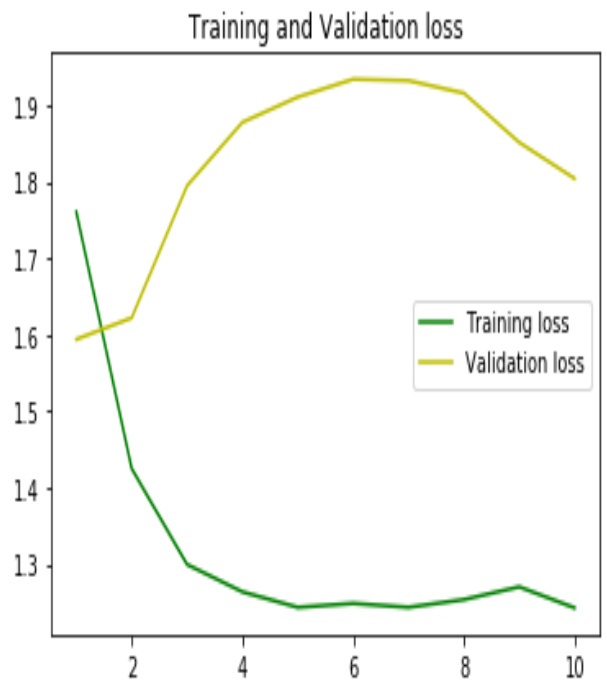


Figure 17 Training and Validation Loss Analysis

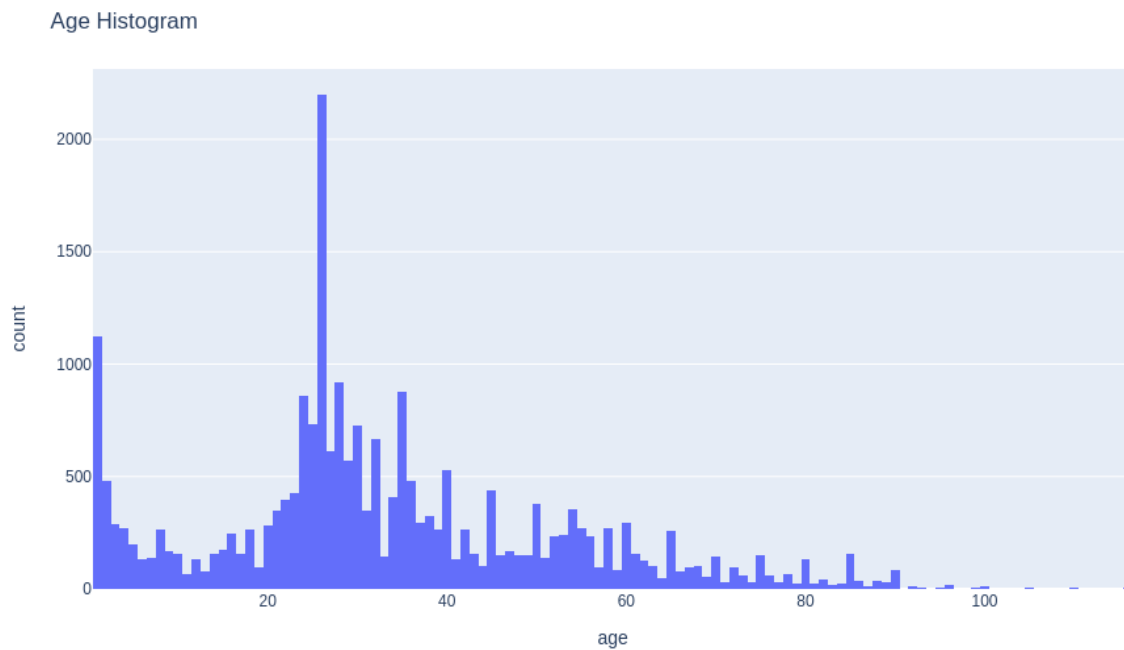


Figure 18 Age Histogram

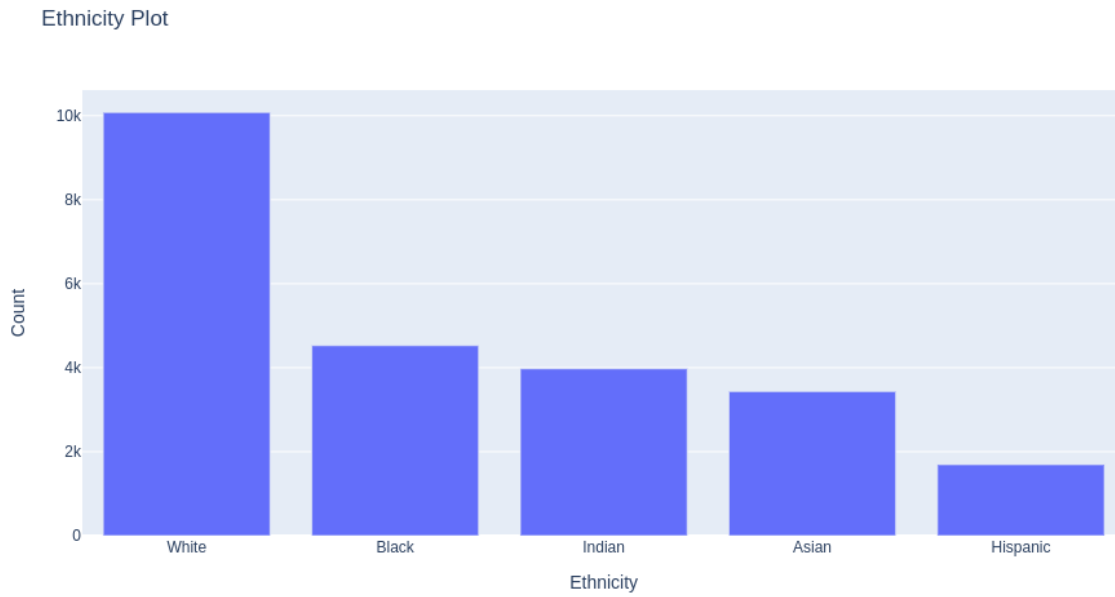


Figure 19 Ethnicity Histogram

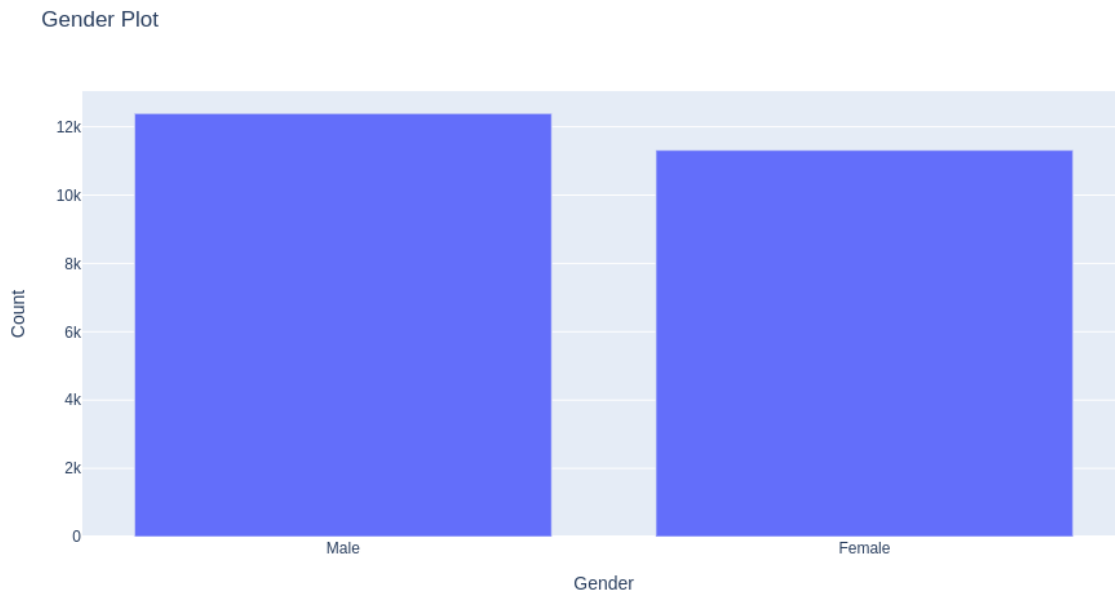


Figure 20 Gender Histogram

6.4 Evaluation result

Table 10 Evaluation summary for the TomNet model

No	Function	LR	Epoch	Data split ratio	Optimizer	Training accuracy	Validation Accuracy	Image Type
1	Softmax	0.001	10	8:2	ADAM	99.51%	99.37%	RGB
2	Softmax	0.01	10	8:2	ADAM	99.72%	99.79%	RGB
3	Softmax	0.1	10	8:2	ADAM	99.88%	99.99%	RGB
4	Sigmoid	0.001	10	8:2	ADAM	99.15%	99.99%	RGB
5	Sigmoid	0.01	10	8:2	ADAM	98.79%	99.44%	RGB
6	Sigmoid	0.1	10	8:2	ADAM	98.12%	98.22%	RGB
7	Softmax	0.001	10	8:2	ADAM	96.97%	84.27%	Grayscale
8	Softmax	0.01	10	8:2	ADAM	97.16%	76.35%	Grayscale
9	Softmax	0.1	10	8:2	ADAM	98.34%	78.75%	Grayscale
10	Sigmoid	0.001	10	8:2	ADAM	96.97%	84.27%	Grayscale
11	Sigmoid	0.01	10	8:2	ADAM	98.55%	81.35%	Grayscale
12	Sigmoid	0.1	10	8:2	ADAM	98.89%	83.23%	Grayscale

The above table summarizes top result found for the evaluation of the TomNet model on two types of datasets which are used for comparison. This dataset types are the grayscale image type which is types of images with only one-color channel the RGB image types which are images contains more than 16.7 million color in one RGB images.

```

In [27]: 1 acc = history.history['accuracy']
          2 val_acc = history.history['val_accuracy']
          3 loss = history.history['loss']
          4 val_loss = history.history['val_loss']
          5 epochs = range(1, len(acc) + 1)
          6 #Train and validation accuracy
          7 plt.plot(epochs, acc, 'g', label='Training accuracy')
          8 plt.plot(epochs, val_acc, 'y', label='Validation accuracy')
          9 plt.title('Training and Validation accuracy')
         10 plt.legend()
         11
         12 plt.figure()
         13 #Train and validation loss
         14 plt.plot(epochs, loss, 'g', label='Training loss')
         15 plt.plot(epochs, val_loss, 'y', label='Validation loss')
         16 plt.title('Training and Validation loss')
         17 plt.legend()
         18 plt.show()

```

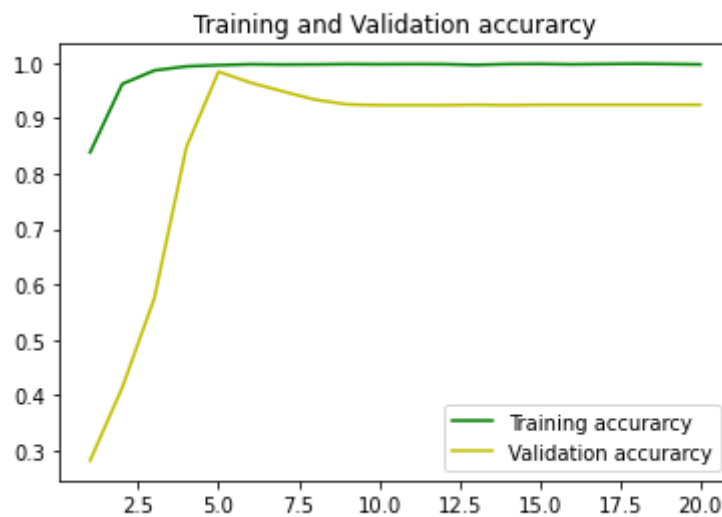


Figure 21 Accuracy Graph of Train and Validation of images

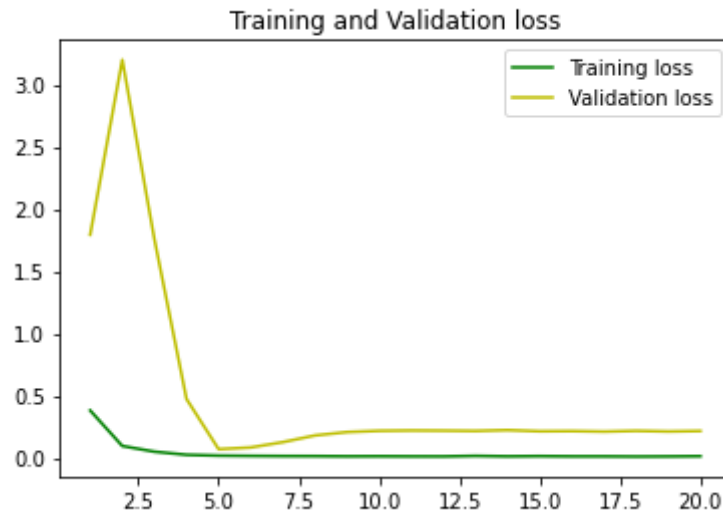


Figure 22 Loss Graph of Train and Validation of images

The loss graph shows that the model loss which starts from around 3% has gradually decreased and tends to almost zero and this is what we are looking for and at the same time if we see the accuracy to 98% and above. In both graphs both the training and validation data have no significant distance between them. There 3 results are evaluated using 20 number of epoch with test ratio 15% and LR of 0.001 which result in an accuracy of 98.34%.

CHAPTER SEVEN: CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this paper Face Recognition and Detection can act as security measures. The authentications can be easily done using this thesis through the classification of individuals Age, Gender, Ethnicity using a convolutional neural network (CNN).

This study's experiment is focused on four common face features. Nineteen thousand one hundred eight (19108) augmented images are created from these original images. The target TomNet is modeled and trained using these images. The model is used to classify and face recognition images into their respective categories. The model categorizes the input image based on what it learned during training. The target classes are the features that were chosen for this study. Age, Gender, Ethnicity, and Pose are the classes. This research finds the more efficient model with a training accuracy of 98.88%. The validation accuracy of this model is 98.34% and the testing accuracy 98.56%. The research result is more efficient and solves the mentioned problem if properly deployed. The images take of format JPEG or PNG extension. The face images used for training and testing are taken both from face databases.

7.2 Future Work

For more than two-decade different approach and algorithms are used for face detection and Recognition. Each study are almost works in terms of accuracy, but no more studies related to face Detection and Recognition is worked accuracy with respect to speed to be recognized. My recommendation for Researcher the parameter of Detection and Recognition is not only Accuracy and Error. The speed of Recognition is also very necessary in terms of timeliness.

The problem with Face Detection and Recognition is no more studies are consideration is when the input data are not accurate. This is mean that if the images in training and testing face have high noise it is difficult to recognize. So further studies are needed to work

REFERENCE

1. D. Arun Kumar, B. Iniyan, M. Ahamed Askar, A. Ajay, R. Ambika,” Face recognition based new generation ATM machine”, 2019 5th International Conference on Advanced Computing & Communication Systems.
2. Suleman Khan, M. Hammad Jaed et.al “Facial Recognition using Convolutional Neural Networks and Implementation on smart Glasses” 2019 international conference on information science and communication technology(ICISCT).
3. Dubois, Antoine “Facial Recognition using deep neural networks” 2018,
4. F. Galton, “Personal identification and description 1” Nature, pp.173-177, 21 June1888.
5. Sir Francis Galton, “Personal identification and description-II”, Nature, pp. 201- 203, 28 June 1888.
6. Jie Wang and Zihao Li, “Research on Face Recognition Based on CNN”, 2018 2nd International symposium on resource exploration and environmental science.
7. Richard Szeliski, Computer Vision: Algorithms and Applications (2010) pp 10-11
8. https://en.wikipedia.org/wiki/Computer_vision Retrieved June 4, 2018.
9. DeepFace:-Closing the gap to human level performance in face recognition.(Conference on Computer Vision and Pattern Recognition (CVPR)
10. T. Ahonen, A. Hadid, and M. Pietikäinen,” Face Recognition with Local Binary Patterns” (CCV 2004, LNCS 3021), pp. 469–481.
11. Weihong Wang, Jie Yang, Jianwei Xiao, Sheng Li, and Dixin Zhou, “Face Recognition Based on Deep Learning”, Zhejiang University of Technology.
12. Sharma S, Karthikeyan Shanmugasundaram2, Sathees Kumar Ramasamy,” FAREC-CNN based efficient Face Recognition Technique using Dlib”, 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)
13. Ni Kadek Ayu Wirdiani, Praba Hridayami, Ni Putu Ayu Widiari, Komang Diva Rismawan, Putu Bagus Candradinatha, I Putu Deva Jayantha,” Face Identification on K-Nearest Neighbor”, 2019 scientific Journal of information
14. Jignesh J. Patoliya, Miral M. Desai, “Face Detection Based ATM Security System Using Embedded Linux Platform”, 2017 2nd International Conference for Convergence in Technology.

15. Faizan Ahmad, Aaima Najam, and Zeeshan Ahmed, "Image based Face Detection and Recognition.
16. Ahmad Tasnim Siddiqui, "Biometrics to control ATM scams: A study", 2014 International Conference on Circuit, Power and Computing Technologies[ICCPCT].
17. M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cognitive Neuroscience, vol. 3, 1991.
18. P.J Phillips, "Support vector machine applied to face recognition." processing system 11, 1999.
19. J. Huang, V. Blanz, and B. Heisele, "Face recognition using Component-Based support vector machine Classification and Morphable models," LNCS 2388, 2002.
20. K. Jonsson, J. Mates, J. Kittler and Y.P. Li, "learning support vectors for face verification and recognition," Fourth IEEE International conference on Automatic face and Gesture Recognition 2000, pp.208-213, Los Alamitos, USA, March 2000.
21. Y. Gao and K.H. Leung, "Face recognition using line edge map" IEEE transaction on pattern analysis and machine intelligence, vol 24, no. 6, June 2002.
22. O. Deniz, M. Castrillon, M. Hernandez, "face recognition using independent component analysis and support vector machines" Pattern recognition letters, vol. 24, pp. 2153-2157, 2003.
23. P.a. G. Adam, "Deep learning a actioner's Approach, Sebastopol, " O'Reilly media, 2017
24. J. L. E. M. A. R. D. O. F. B. R. H. S. A. N. a. N. D. R. Miikkulainen, "Evolving deep neural networks, "in Artificial intelligence in the Age of Neural Networks and Brain Computing," *Elsevier*, pp. 293-31, 2019.
25. Josh A. Black, M. Gargsha, K. Kahol, P. Kuchi, Sethuraman Panchanathan" A Framework performance evaluation of face recognition algorithm," in proceeding of the international Conference on ITCOM. Internet Multimedia System II, 2002.
26. The At&T database Available on <http://www.uk.research.att.com/facedatabase.html>
27. The Yale database available on <http://cvc.yale.edu/>
28. The MIT face database available on <http://whitehapel.media.mit.edu/pub/images/>
29. www.atmmarketplace.com
30. David Habrman, "Face Recognition with Preprocessing and Neural Network" Linkoping, Sweden, 2016.

31. W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: a literature Survey", Technical Report, Univ. of Maryland, 2000.
32. R. Mauro and M. KUbovy, "Caricature and face Recognition," Memory and Cognition, Vol.20, No.4, pp. 433-441, 1992.
33. Josh D. Woodward, Jr., Christopher Horn, Julius Gatune, and Aryn Thomas., "bio-metrics A Look at Facial Recognition", Santa Monica, CA: RAND, 2001.
34. Mansoor Roomi, Dr.M.Parisa Beham, "A Review of Face Recognition Methods", in international journal , April 2013.
35. M. A. Turk, and A. P. pentland, "Face Recognition using Eigenfaces", pp. 586-591, IEEE, 1991.
36. Annie Xin Guan, Harold H. Szu, "A local Face Statistics Recognition Methodology beyond ICA and/or PCA, 1999.
37. <https://www.opencv.org>
38. V.p. Vishwakarama, S. Pandey and M.N. Gupta, "A Novel Approach for Face Recognition Using DCT Re-scaling for Illumination Normalization" 15th International Conference on Advanced Computing and Communication.
39. Navpeet Kaur Review of Face Recognition System Using MATLAB international journal of computer science trends and technology(IJCST), volume 4 Issue 3, May- Jun 2016, pp 20-33.
40. N. T. Deshpande, Dr. S. Ravishankar, Face Detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN, Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1173-1189.
41. https://www.tutorialspoint.com/dip/histogram_equalization.htm
42. Lusekelo Kibona, Face Recognition as a Biometric Security for Secondary Password for ATM User. A Comprehensive Review,2015.
43. Ian.T.Young, Lucas J Van Vliet, Fundamentals of Image processing, Article July 2004.
44. C.Francois," Deep Learning with Python" New York: Manning Publication, 2017.
45. N. T. Deshpande, Dr. S. Ravishankar , Face Detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN, Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1173-1189
46. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: proceeding of CVPR 1:511-518.

47. Mengjia Yan, Mengao Zhao, Zining Xu et.al,” VarGFaceNet: An Efficient Variable Group Convolutional Neural network for Lightweight face Recognition”

APPENDIXES

Appendix I: Experiment of TomNet Model

```
In [1]: 1 import numpy as np
        2 import tensorflow.keras as keras
        3 from keras import backend as k
        4 from keras.regularizers import l1
        5 from tensorflow.keras.layers import Dense, Dropout, GlobalAveragePooling2D,
        6 from tensorflow.keras.optimizers import Adam
        7 from keras.metrics import categorical_crossentropy, categorical_accuracy
        8 from keras.preprocessing.image import ImageDataGenerator
        9 from keras import models, regularizers
       10 from keras.applications import imagenet_utils
       11 from sklearn.metrics import confusion_matrix
       12 import itertools
       13 import matplotlib.pyplot as plt
       14 from keras.callbacks import EarlyStopping, ModelCheckpoint, TensorBoard
       15 import tensorflow as tf
       16 import os
       17 from tensorflow.keras.models import Sequential
       18 from tensorflow.keras.layers import Flatten, Conv2D
```

Appendix II: Code for Reading Images

```
In [2]: 1 train_data_dir = 'images1/train_data'
        2 validation_data_dir = 'images1/validation_data'
        3 test_data_dir = 'images1/test_data'
```

Appendix III: Code for Image Augmentation

```
In [3]: 1 # Let's use some data augmentaiton
2 train_datagen = ImageDataGenerator(
3     rescale=1. / 255,
4     rotation_range=30,
5     width_shift_range=0.3,
6     height_shift_range=0.3,
7     horizontal_flip=True,
8     fill_mode='nearest')
9
10 validation_datagen = ImageDataGenerator(rescale=1. / 255)
11
12 train_generator = train_datagen.flow_from_directory(
13     train_data_dir,
14     target_size=(img_rows, img_cols),
15     batch_size=batch_size,
16     class_mode='categorical',
17     shuffle=True)
18
19 validation_generator = validation_datagen.flow_from_directory(
20     validation_data_dir,
21     target_size=(img_rows, img_cols),
22     batch_size=batch_size,
23     class_mode='categorical',
24     shuffle=False)
```

Appendix IV: Model testing sample

```
In [14]: 1 import plotly.graph_objects as go
2 def plot_data(rows, cols, lower_value, upper_value):
3     fig = plt.figure(figsize=(cols*3,rows*4))
4     for i in range(1, cols*rows + 1):
5         k = np.random.randint(lower_value,upper_value)
6         fig.add_subplot(rows, cols, i) # adding sub plot
7         gender= gender_values_to_labels[df.gender[k]]
8         ethnicity =eth_values_to_labels[df.ethnicity[k]]
9         age = df.age[k]
10        im = df.pixels[k]
11        plt.imshow(im, cmap='gray')
12        plt.axis('off')
13        plt.title(f'Gender:{gender}\nAge:{age}\nEthnicity:{ethnicity}')
```

