AN ENGINEERING MAJOR PROJECT FINAL REPORT
ON
REAL-TIME AGE AND GENDER DETECTION

Submitted By

Anil Ghimire	190101
Mahesh Bhatt	190111
Devraj Chaudhary	190144
Sahil Poudel	190147
Sandip Subedi	190148

Under the Supervision of
Ranjan Raj Aryal

Submitted To
The Department of Information and Communications Technology
in partial fulfillment of requirement for the degree of Bachelor of
Engineering in Information Technology

Cosmos College of Management & Technology
(Affiliated to Pokhara University)
Tutepani, Lalitpur, Nepal

July 2024

AN ENGINEERING MAJOR PROJECT FINAL REPORT
ON
REAL-TIME AGE AND GENDER DETECTION

Submitted By

Anil Ghimire	190101
Mahesh Bhatt	190111
Devraj Chaudhary	190144
Sahil Poudel	190147
Sandip Subedi	190148

Submitted To

The Department of Information and Communications Technology
in partial fulfillment of requirement for the degree of Bachelor of
Engineering in Information Technology

Cosmos College of Management & Technology
(Affiliated to Pokhara University)
Tutepani, Lalitpur, Nepal

July 2024

COPYRIGHT

CERTIFICATE

The undersigned certify that they have read recommended to the Department Information and Communications Technology, a final year project work entitled “Real-Time Age And Gender Detection” submitted by Anil Ghimire (190101), Mahesh Bhatt (190111), Devraj Chaudhary (190144), Sahil Poudel (190147), Sandip Subedi (190148) in partial fulfillment of the requirements for the degree of Bachelor of Engineering in Information Technology.

Ranjan Raj Aryal
(Project Supervisor)
Department of Information and Communications Technology
Cosmos College of Management and Technology

Santosh Giri
(External Examiner)
Assistant Professor
Pulchowk Campus

Bibek Rupakheti
Head of the Department
Department of Information and Communications Technology
Cosmos College of Management and Technology

Acknowledgments

We would like to express our sincere special thanks to our Project Supervisor Er.Ranjan Raj Aryal, Head Department of IT and Computer Engineering Er.Bibek Rupakheti who had given their valuable time and given us a chance to learn something, also we would like to thank our Asst. Professor and Deputy Head Department of Information and Communication Technology Er.Chirangibi Pandey for his support and help despite having a busy schedule for their great guidelines. Also we would thank all those people who gave us the golden opportunity to work in this wonderful project on the topic Real-Time Age And Gender Detection, which we hope would help us in doing a lot of Research and let us to know about so many new things about existing technology and programming language, we would be really thankful to them.

Secondly, we would also like to thank my parents and friends who helped us a lot in finalizing this project topic.

Abstract

Automatic age and gender classification has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. However, performance of existing methods on real-world images is still significantly lacking, especially when compared to the tremendous leaps in performance recently reported for the related task of face recognition. Thus, this report is prepared to show that by learning representations through the use of deep-convolutional neural networks (DCNN), a significant increase in performance can be obtained on these tasks. The two-level CNN architecture includes feature extraction and classification itself. The feature extraction extracts feature corresponding to age and gender, while the classification classifies the face images to the correct age group and gender. This is done by using deep learning, OpenCV which is capable of processing the real-time frames which is given as input and the determined age and gender as output based on the evaluation of method on the recent UTKFace (dataset) for gender and age estimation. The evaluation method includes classification rate, precision, and recall using UTKface dataset and real-world images to exhibit excellent performance by achieving good prediction results and computation time with validation accuracy 88 % on gender detection and 7.21 mean absolute error for age detection.

Contents

List of Figures

List of Abbreviations

AAM	Active Appearance Model
ANN	Artificial Neural Network
BIF	Biologically Inspired Features
CNN	Convolutional Neural Network
CCA	Canonical Correlation Analysis
CCTV	Close Circuit Television
IMDb	Internet Movie Database
LBP	Local Binary Patterns
OpenCV	Open Source Computer Vision Library
SVM	Support Vector Machine
SVR	Support Vector Regression
VOG	Visual Geometry Group

1. Introduction

1.1 Background

Facial analysis has gained much recognition in the computer vision community in the recent past. Human's face contains features that determine identity, age, gender, emotions, and the ethnicity of people. Among these features, age and gender classification can be especially helpful in several real-world applications including security and video surveillance, electronic customer relationship management, biometrics, electronic vending machines, human-computer interaction, entertainment, cosmetology, and forensic art. A lot of research has been done using deep learning methods such as ANN, CNN to determine age and gender estimation. Fundamental facial consideration features are eyebrows, mouth, nose and eyes. An architecture based on the convolution Neural network (CNN) is proposed here for age and gender classification. This is one of the well-known deep artificial neural networks. Convolutional Neural Network based design models are broadly utilized in classification task because of their remarkable execution in facial investigation. The Convolutional Neural Network includes feature extraction which extracts features corresponding to age and gender. Furthermore CNN includes feature classification which classifies facial images into the correct age and determines the gender. In current world, works in age and gender classification is showing encouraging signs of progress in deep learning and CNN, therefore end-toend deep learning-based classification model is proposed here that predicts age group and gender of unfiltered facial images. The age and gender classifications task as a classification problem is formulated in which the CNN model learns to predict the age and gender from a face.

1.2 Problem statements

CCTV footages can show the criminal activities but can't deduce the culprit easily. Analysis and prediction of the customer's need varies as per the age, so marketing strategy for different age groups on different platform is a hurdle. Quick biometrics tests could simplify the efforts needed to save one's life. Use of Alcohol among lots of teens has been a major issue due to easy access to vending machines that provides the alcoholic beverages without being able to consider the possibility of underage kids taking it. Several such issues exist at present, and all these and many others may be avoided for good. This project emphasis on eliminating all these issues. It can help enhance the marketing policy, reduce the time needed to find culprit, diagnosis the health issue without much delay and so on. Simply detecting the age and gender can assist in numerous problems and not to mention, numerous fields.

1.3 Objectives

Following are the major objectives:

- To detect the face/s from real time video
- To determine the age and gender of the detected face/s

1.4 Scope

Age and gender detection and classification has its scope in numerous field. It can be used for forensic testing, security and video surveillance, human-computer interaction, cosmetology, electronic vending machines, marketing purposes and so on. Major applications of the project includes:

- Easy detection of age and gender in forensic or biometrics helps deduce the conclusion faster
- Age and gender determination can reduce the effort to search the culprit, hence helpful for video surveillance and security at the same time
- Useful for marketing proposes i.e. showing ads on different platforms as per the age and gender, surely would be fruitful
- Can be used to automate the access to adult content sites or any other platforms having age-limit criteria
- Can be used to restrict access of alcohol from vending machines
- Useful for editing apps or software related to cosmetology

2. Literature Review

Facial analysis has gained much recognition in the computer vision community in the recent past due to its enormous application and possibilities. Human's face contains features that determine age, gender, emotions, ethnicity and identity of people. Among these features, age and gender classification can be especially helpful in several real-world applications including security and video surveillance, electronic customer relationship management, biometrics, electronic vending machines, humancomputer interaction, entertainment, cosmetology, and forensic art. However, several issues in age and gender classification are still open enigma. Age and gender predictions of unfiltered real-life faces are yet to meet the requirements of commercial and real-world applications in spite of the scrutiny computer vision community keeps making with the continuous amelioration of the new techniques that improves the state of the art [1][2][3].

Over the past years, a lot of methods have been proposed to solve the classification issues. Many of those methods are handcrafted which perform unsatisfactorily on the age and gender predictions of unconstrained in-the-wild images [2][4]. These conventional hand-engineered methods relied on the differences in dimensions of facial features and face descriptors [5][6][7] which do not have the ability to handle the varying degrees of variation observed in these challenging unconstrained imaging conditions. The images in these categories have some variations in appearance, noise, pose, and lighting which may affect the ability of those manually designed computer vision methods to accurately classify the age and gender of the images. Recently, deep learning-based methods [8][9] have shown encouraging performance in this field especially on the age and gender classification of unfiltered face images. In light of the current works in age and gender classification and encouraging signs of progress in deep learning and CNN, a deep learning-based classification model that predicts age group and gender of unfiltered facial images has been proposed in this report. The age and gender classifications task has been formulated as a classification problem in which the CNN model learns to predict the age and gender from a face image.

Almost all of the early methods in age and gender classifications were handcrafted, focusing on manually engineering the facial features from the face and mainly providing a study on constrained images that were taken from controlled imaging conditions. To mention a few, in 1999, Kwon and Lobo [10] developed the very first method for age estimation focusing on geometric features of the face that determined the ratios among different dimensions of facial features. These geometric features separated babies from adult successfully but were incapable of distinguishing between young adult and senior adult. Hence, in 2004, Lanitis et al. [11] proposed an Active Appearance Model (AAM) based method that included both the geometric and texture features, for the estimation task. This method was not suitable for the unconstrained imaging conditions attributed to real-world face images which have different degrees of variations in illumination, expression, poses, and so forth. From 2007, most of the approaches employed manually designed features for the estimation task: Gabor [5], Spatially Flexible Patches (SFP) [6], Local Binary Patterns (LBP) [12], and Biologically Inspired Features (BIF) [13]. Classification methods in [3][14] used Support Vector Machine (SVM) based methods for age and gender classification. Linear regression [7][15], Support Vector Regression (SVR) [16], Canonical Correlation Analysis (CCA) [17], and Partial Least Squares (PLS) [18] are the common regression methods for age and gender predictions. Dileep and Danti [19] also proposed an approach that used feed-forward propagation neural networks and 3-sigma control limits approach that classified people's age into children, middleaged adults, and old-aged adults. However, all of these methods were only suitable and effective on constrained imaging conditions; they couldn't handle the unconstrained nature of the real-world images and therefore, couldn't be relied on to achieve respectable performance on the in-the-wild images which are common in practical applications [3].

More recently, an expanding number of researchers started to use CNN for age and gender classification. It could classify the age and gender of unfiltered face images relying on its good feature extraction technique [8][9][20]. Availability of sufficiently large data for training and high-end computer machines also helped in the adoption of the deep CNN methods for the classification task. CNN model can learn, compact and discriminative facial features, especially when the volume of training images is sufficiently large, to obtain the relevant information needed for the two classifications. For example, in 2015, Levi et al. [4] proposed a CNN based model, comprising of five layers, three convolutional and two fully connected layers, to predict the age of real-world face images. The model included center-crop and oversampling method, to handle the small misalignment in unconstrained images. Yi et al. [21], in their paper, applied an end-to-end multitask CNN system that learns a deeper structure and the parameters needed, to solve the age, gender, and ethnicity classification task. In [22], the authors investigated a pre-trained deep VGG-Face CNN approach, for automatic age estimation from real-world face images. The CNN based model consists of eleven layers, including eight convolutional and three fully connected layers.The authors in [1] also proposed a novel CNN based method, for age group and gender estimation: Residual Networks of Residual Networks (RoR). The model includes an RoR architecture, which was pretrained on gender and weighted loss layer and then on ImageNet dataset, and finally it was fine-tuned on IMDb-WIKI-101 dataset. Ranjan et al. in [23] presented a model that simultaneously solved a set of face analysis tasks, using a single CNN. The end-to-end solution is a novel multitask learning CNN framework, which shared the parameters from lower layers of CNN among all the tasks for gender recognition, age estimation, etc. In [2], the authors proposed a CNN solution for age estimation, from a single face image. The CNN based solution included a robust face alignment phase that prepared and preprocessed the face images before being fed to the designed model. The authors also collected large-scale face images, with age and gender label: IMDb-WIKI dataset. In 2018, Liu et al. [24] developed a CNN based model that employed a multiclass focal loss function. The age estimation model was validated on Adience benchmark for performance accuracy, and it achieved a comparable result with stateof-the-art methods.

Unfortunately, some of these methods mentioned above have been verified effectively on constrained imaging conditions; few studied the unconstrained imaging conditions. Still, it is a challenging task to classify unconstrained faces with large variations in illumination, viewpoint, nonfrontal, etc. Here, those issues has been addressed by designing a robust image preprocessing algorithm, pretraining the model on large-scale facial aging benchmarks with noisy age and gender labels, and regularize the CNN parameters with self-designed CNN framework.

3. Methodology

In order to classify the unconstrained faces, image preprocessing stage is required that preprocess and prepare the face images before they are input into the proposed network. Therefore, to accomplish the whole process the solution is divided into three major steps: image preprocessing, features learning, and classification.

Image preprocessing included resizing of image and grey scale conversion. Feature learning included the use of convolution layers which applied a set of learnable filters to the input image to extract relevant features. Classification included probability distribution to predict the relevant class.

3.1 Block Diagram

A block diagram is a visual representation of a system that uses simple, labeled blocks that represent single or multiple items, entities or concepts, connected by lines to show relationships between them.

The block diagram representing the methodology for our project is shown below:

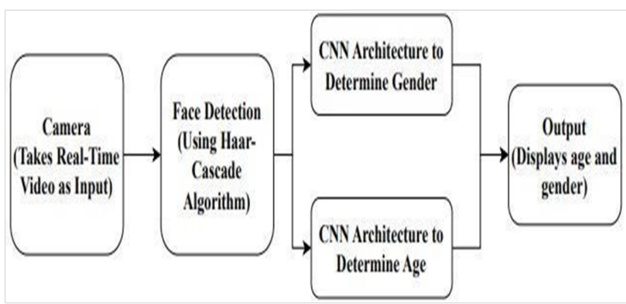


Figure 3.1: Block Diagram

The camera is used as the input source through which a real-time video is taken for the system. The video is further processed by the system to detect the face, determine the age and gender and classify them.

When a frame/video is input, the Haar-Cascade algorithm first detects for faces in each frame. Once it find faces in the frame, the face is fed to CNN architecture used to determine gender which consists of two labels; essentially Male and Female and gender is detected. Again, for age detection, the face detected using Haar-cascade is fed to CNN architecture used to determine age and here age is determined using regression. The determined age may fall between 0-116 years. Finally, the result is displayed on the frame containing the gender and age using OpenCV. The resulting frame consists of the square box around the face/s with the estimated gender and the age.

3.2 Algorithm

Algorithm is a process or set of rules to be followed in calculations or other problemsolving operations. Four Algorithms (face detection, gender detection, age classification and testing) followed for project's accomplishment are explained below:

3.2.1 Face Detection: Haar-Cascade Classifier

The Haar-cascade algorithm is a machine learning-based approach for object detection, which was originally proposed by Viola and Jones in 2001 for detecting faces in images. The algorithm works by using a set of Haar-like features and a cascade classifier to identify objects of interest. The algorithm is:

Step 1: Collect positive (image that contain face/s) and negative (image that don't contain face/s) samples

Step 2: Extract Haar-like features (rectangular patterns that can detect edges, lines and corners in an image) from the samples

Step 3: Train a classifier using the AdaBoost algorithm (algorithm works by iteratively selecting the most informative features and training weak classifiers on them, weak classifiers are combined to form strong classifier that can accurately detect faces)

Step 4: Create a cascade (series of stages) of weak classifiers

Step 5: Apply the cascade to each region of the image to detect faces

Step 6: Perform post-processing to remove false positives and refine the locations of the detected faces

3.2.2 Gender Detection

Once the face is detected using above algorithm, next step is to identify the gender from that face. For that, the algorithm used is listed below:

Step 1: Detect faces in the input image using the Haar-cascade algorithm

Step 2: Preprocess the detected faces by resizing and gray scaling them to a fixed size

Step 3: Feed the preprocessed faces into a trained model

Step 4: The model extracts features from the input image and make a prediction on the gender

Step 5: The output of the model will be a probability distribution over the possible classes (male or female)

Step 6: The class with the highest probability will be chosen as the predicted gender

3.2.3 Age Detection

Once the gender is detected, age classification is done and is classified in the range 0116 years. The algorithm is:

Step 1: Detect faces in the input image using the Haar-cascade algorithm

Step 2: Preprocess the detected faces by resizing and gray scaling them to a fixed size

Step 3: Feed the preprocessed faces into a trained model for regression

Step 4: The model extracts features from the input image and output a continuous value representing the estimated age from the detected face

3.2.4 Testing

Algorithm for testing is numbered below: Step 1: Detect faces in the input image using the Haar-cascade algorithm

Step 2: Preprocess the detected faces by resizing and gray scaling them to a fixed size

Step 3: Feed the processed faces data into a trained model

Step 4: Load faces image from training directory for prediction

Step 5: Retrieve matched image's age from database

Step 6: Display Result

3.3 System Flow Chart

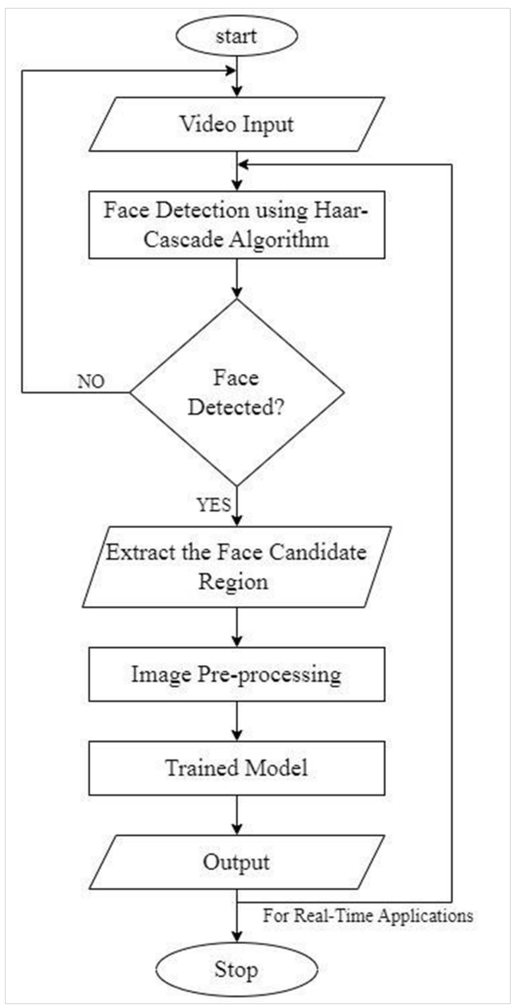


Figure 3.2: System Flow Chart

3.4 Model Implementation

3.4.1 Dataset

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. Hence, this dataset has been utilized to accomplish the project. The labels of each face image is embedded in the file name, formatted like: [age]_[gender]_[race]_[date&time].jpg

• [age] is an integer from 0 to 116, indicating the age

• [gender] is either 0 (male) or 1 (female)

• [race] is an integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern)

• [date&time] is in the format of yyyymmddHHMMSSFFF, showing the date and time an image was collected to UTKFace

Among 23,078 images on UTKFace dataset, 35% have been used for testing and the remaining 65% for training the model.

3.4.2 CNN

A CNN is a type of feedforward network structure that is formed by multiple layers of convolutional filters alternated with subsampling filters followed by fully connected layers. They are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images. Their applications range from image and video recognition, image classification, medical image analysis, computer vision and natural language processing. The term 'Convolution" in CNN denotes the mathematical function of convolution which is a special kind of linear operation wherein two functions are multiplied to produce a third function which expresses how the shape of one function is modified by the other. In simple terms, two images which can be represented as matrices are multiplied to give an output that is used to extract features from the image.

CNN Architecture

There are two main parts to a CNN architecture:

- A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.

- A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

The network of feature extraction consists of many pairs of convolutional or pooling layers. CNN model of feature extraction aims to reduce the number of features present in a dataset. It creates new features which summarizes the existing features contained in an original set of features.
A fully connected layer is comprised of flatten and dense layers. Flatten layers takes the 3D output tensor from the previous layer and converts it into a 1D array, which is then fed into the dense layer. The dense layer then map the flattened feature vector to the target output class using a set of learnable weights and biases.
The CNN Architecture for the project is explained hereby :

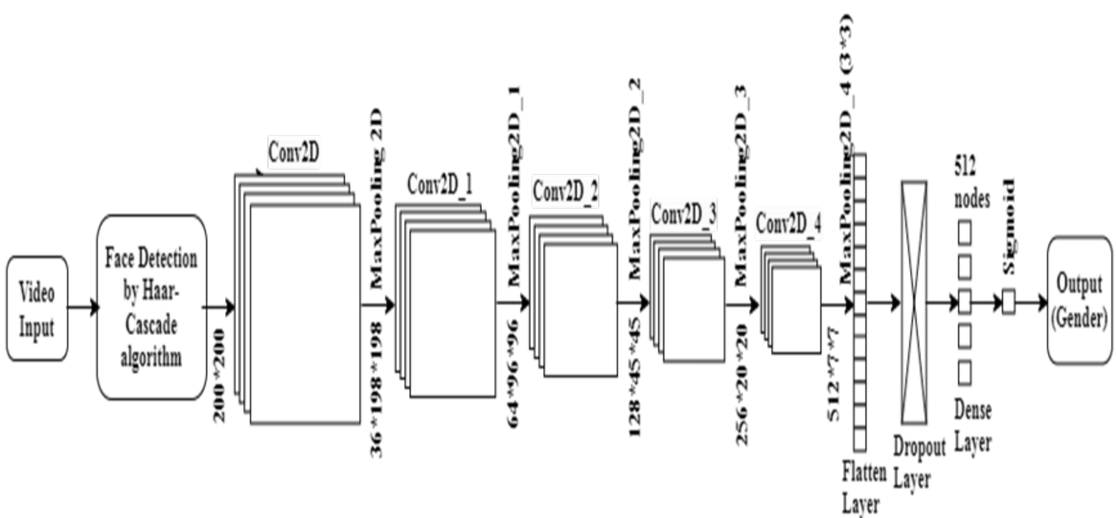


Figure 3.3: CNN Architecture for Gender Detection

Images are initially rescaled to 200*200 pixels and then sent to the convolution layers. Following that, the five convolutional layers are defined as follows:

- The first convolutional layer applies 36 filters of size 36*198*198 pixels to the input, followed by a rectified linear operator (ReLU), a max pooling layer that takes the maximum value of 3*3 regions with two-pixel steps, and a local response normalization layer

- The second convolutional layer, which contains 64 filters of size 64*96*96 pixels, processes the max layer's 64*47*47 output. The same hyper parameters as before are used for ReLU, a max pooling layer, and a local response normalization layer

- The third convolutional layer applies a set of 128 filters of size 128*45*45 pixels and max layer applied 128*22*22, followed by ReLU and a max pooling layer

- The fourth convolution layer, which has 256 filter of size 256*20*20 pixels and max pooling layer is 256*9*9, is followed by a ReLU and a dropout layer

- A fifth convolution layer , which has 512 filter of size 512*7*7 pixels and max pooling layer 512*3*3 pixels, followed by a ReLU ,a flatten and a dropout layer

- Finally, a dense layer with 512 neurons is applied followed by cross entropy for prediction of gender

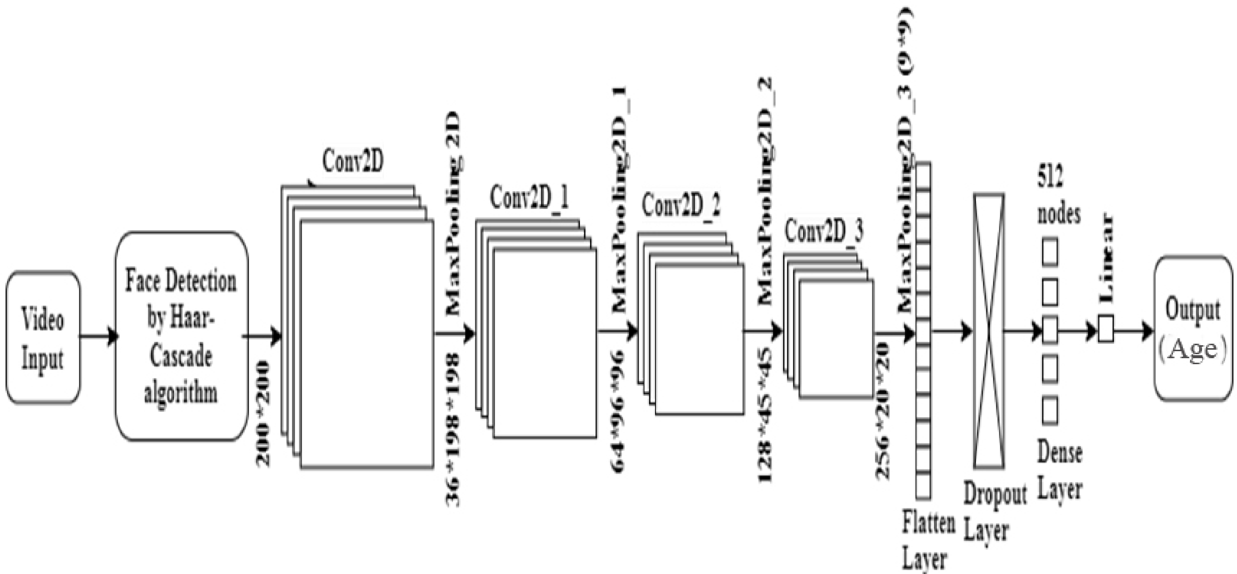


Figure 3.4: CNN Architecture for Age Detection

Images are initially rescaled to 200*200 pixels and then sent to the convolution layers. Following that, the four convolutional layers are defined as follows:

- The first convolutional layer applies 36 filters of size 36*198*198 pixels to the input, followed by a rectified linear operator (ReLU), a max pooling layer that takes the maximum value of 3*3 regions with two-pixel steps, and a local response normalization layer

- The second convolutional layer, which contains 64 filters of size 64*96*96 pixels, processes the max layer's 64*47*47 output. The same hyper parameters as before are used for ReLU, a max pooling layer

- The third convolutional layer applies a set of 128 filters of size 128*45*45 pixels and max layer applied 128*22*22, followed by ReLU and a max pooling layer

- A four convolution layer , which has 256 filter of size 256*20*20 pixels and max pooling layer 256*9*9 pixels, followed by a ReLU, a flatten and a dropout layer

- Finally, a dense layer with 512 neurons is applied followed by a linear function for age classification

3.4.3 Summary of Model Layer

CNN architecture for Gender Detection is comprised of 5 convolutional layers with a fully connected layers, summarized below:

- An input 2D convolutional layer(with 36 filters) paired with a 2D MaxPooling layer
- 4 pairs of 2D convolutional layers with 64,128,256 & 512 filters respectively paired again with 2D MaxPooling layers
- 1 Flatten layer and then 1 Dropout Layer
- 1 Dense layer with 512 nodes and finally
- An output Dense layer with 2 nodes which are essentially, labels; male or female

Layer (type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None, 198, 198, 36)	1,008
max_pooling2d (MaxPooling2D)	(None, 98, 98, 36)	0
conv2d_1 (Conv2D)	(None, 96, 96, 64)	20,800
max_pooling2d_1 (MaxPooling2D)	(None, 47, 47, 64)	0
conv2d_2 (Conv2D)	(None, 45, 45, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 22, 22, 128)	0
conv2d_3 (Conv2D)	(None, 20, 20, 256)	295,168
max_pooling2d_3 (MaxPooling2D)	(None, 9, 9, 256)	0
conv2d_4 (Conv2D)	(None, 7, 7, 512)	1,180,160
max_pooling2d_4 (MaxPooling2D)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dropout (Dropout)	(None, 4608)	0
dense (Dense)	(None, 512)	2,359,808
gender (Dense)	(None, 1)	513

Total params: 11,793,941 (44.99 MB)
Trainable params: 3,931,313 (15.00 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 7,862,628 (29.99 MB)
None

Figure 3.5: Model Layer for Gender Detection

CNN architecture for Age Detection is comprised of 4 convolutional layers with a fully connected layers, summarized below:

- An input 2D convolutional layer(with 36 filters) paired with a 2D MaxPooling layer
- 3 pairs of 2D convolutional layers with 64,128 & 512 filters respectively paired again with 2D MaxPooling layers
- 1 Flatten layer and then 1 Dropout layer
- 1 Dense layer with 512 nodes and finally
- 1 output Dense layer with nodes that specify age

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 198, 36)	1,008
max_pooling2d (MaxPooling2D)	(None, 98, 98, 36)	0
conv2d_1 (Conv2D)	(None, 96, 96, 64)	20,800
max_pooling2d_1 (MaxPooling2D)	(None, 47, 47, 64)	0
conv2d_2 (Conv2D)	(None, 45, 45, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 22, 22, 128)	0
conv2d_3 (Conv2D)	(None, 20, 20, 512)	590,336
max_pooling2d_3 (MaxPooling2D)	(None, 9, 9, 512)	0
flatten (Flatten)	(None, 41472)	0
dropout (Dropout)	(None, 41472)	0
dense (Dense)	(None, 512)	21,234,176
age (Dense)	(None, 1)	513

Total params: 21,920,689 (83.62 MB)

Trainable params: 21,920,689 (83.62 MB)

Non-trainable params: 0 (0.00 B)

None

Figure 3.6: Model Layer for Age Detection

3.4.4 Rectified Linear Unit (ReLU)

The Rectified Linear activation function is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.
ReLU is a non-linear activation function that we used in multi-layer neural networks or deep neural networks. This function can be represented as:

$$f(x) = max(0, x) \tag{3.1}$$

Where x = an input value According to equation 1, the output of ReLU is the maximum value between zero and the input value. An output is equal to zero when the input value is negative and the input value when the input is positive. Thus, we can rewrite equation 1 as follows:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \tag{3.2}$$

Where x = an input value

3.4.5 Adam Optimiser

Adam is an adaptive learning rate optimization algorithm that's been designed specifically for training deep neural networks. The name is derived from adaptive moment estimation. The optimizer is called Adam because it uses estimations of the first and second moments of the gradient to adapt the learning rate for each weight of the neural network. This algorithm is used to accelerate the gradient descent algorithm by taking into consideration the 'exponentially weighted average' of the gradients. Using averages makes the algorithm converge towards the minima in a faster pace. Adaptive Gradient Algorithm (AdaGrad) that maintains a per-parameter learning rate that improves performance on problems with sparse gradients (e.g. natural language and computer vision problems).

3.4.6 Mean Squared Error (MSE)

The Mean Squared Error measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss. Mean square error is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a function. It does this by taking the distances from the points to the regression line (these distances are the "errors") and squaring them. The squaring is necessary to remove any negative signs. It also gives more weight to larger differences.

Mathematically,

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.3}$$

To calculate the mean squared error from a set of X and Y values, first find the regression line and insert the X values into the linear regression equation to find the new Y values. Subtract the new Y value from the original to get the error and then square the errors. Add up the errors and find the mean.

3.4.7 Sigmoid Function

The sigmoid function is a mathematical function that maps any input to a value between 0 and 1. It is often used in machine learning for binary classification tasks such as gender detection.

Here, the sigmoid function is used to predict the probability that a given image belong to a particular gender (e.g., male or female). The sigmoid function takes in the output of the model's final layer, which is typically a weighted sum of the input features, and produces a value between 0 and 1.

If the sigmoid output is closer to 0, the model predicts that the input belongs to the negative class (e.g., male). If the sigmoid output is closer to 1, the model predicts that the input belongs to the positive class (e.g., female).

The decision boundary is adjusted by changing the threshold value to determine whether a sigmoid output is classified as positive or negative.

3.4.8 Linear Function

A linear function is a mathematical equation that represents a straight line on a graph. Here, linear function is used to predict a person's age based on certain input features, such as facial features extracted from an image. The linear function takes the form:

$$age = m * x + b \tag{3.4}$$

Where, age is the predicted age, x is the input feature, m is the slope of the line, and b is the y-intercept. The slope and y-intercept are learned during the training process using a dataset of input features and corresponding age labels. Once the slope and y intercept have been learned, the model uses them to predict the age of new inputs based on their input features.

3.5 Software Process Model

Incremental Model is a process of software development where requirements divided into multiple standalone modules of the software development cycle. In this model, each module goes through the requirements, design, implementation and testing phases. Every subsequent release of the module adds function to the previous release. The process continues until the complete system achieved. The various phases of incremental model are as follows:

- Requirement analysis: In the first phase of the incremental model, the product analysis expertise identifies the requirements. And the system functional requirements are understood by the requirement analysis team. To develop the software under the incremental model, this phase performs a crucial role.

- Design and Development: In this phase of the Incremental model of SDLC, the design of the system functionality and the development method are finished with success. When software develops new practicality, the incremental model uses style and development phase.

- Testing: In the incremental model, the testing phase checks the performance of each existing function as well as additional functionality. In the testing phase, the various methods are used to test the behavior of each task.

- Implementation: Implementation phase enables the coding phase of the development system. It involves the final coding that design in the designing and development phase and tests the functionality in the testing phase. After completion of this phase, the number of the product working is enhanced and upgraded up to the final system product This model is used When the requirements are superior, A project has a lengthy development schedule, When Software team are not very well skilled or trained, and When the customer demands a quick release of the product.

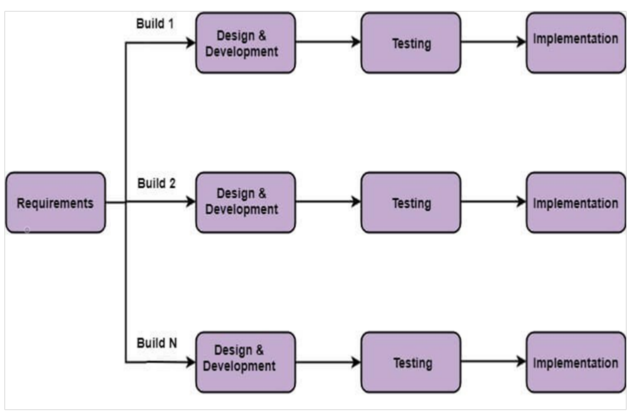


Figure 3.7: Incremental Software Process Model

4. Performance Analysis Methodology

4.1 Model Training and Testing

4.1.1 Training Gender Model

25 epochs were performed for model fitting that are shown below:

```
Epoch 1/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 307s 632ms/step - accuracy: 0.6307 - loss: 4.5897 - val_accuracy: 0.7323 - val_lo
ss: 0.5469
Epoch 2/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 315s 654ms/step - accuracy: 0.7618 - loss: 0.5019 - val_accuracy: 0.8020 - val_lo
ss: 0.4461
Epoch 3/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 308s 639ms/step - accuracy: 0.8088 - loss: 0.4103 - val_accuracy: 0.8460 - val_lo
ss: 0.3584
Epoch 4/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 312s 648ms/step - accuracy: 0.8394 - loss: 0.3572 - val_accuracy: 0.8448 - val_lo
ss: 0.3406
Epoch 5/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 320s 665ms/step - accuracy: 0.8519 - loss: 0.3341 - val_accuracy: 0.8472 - val_lo
```

```
Epoch 23/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 317s 657ms/step - accuracy: 0.9133 - loss: 0.1947 - val_accuracy: 0.8832 - val_lo
ss: 0.2845
Epoch 24/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 304s 631ms/step - accuracy: 0.9149 - loss: 0.2039 - val_accuracy: 0.8823 - val_lo
ss: 0.2742
Epoch 25/25
482/482 ━━━━━━━━━━━━━━━━━━━━ 291s 604ms/step - accuracy: 0.9119 - loss: 0.2024 - val_accuracy: 0.8865 - val_lo
ss: 0.2799
```

Figure 4.1: Model Training process for Gender Detection

- On first epoch, the loss was 4.5897 with accuracy 63% and 0.5469 validation loss with 73% validation accuracy.
- Coming to the 25th epoch, the loss was 0.2024 with 91% accuracy and 0.2799 validation loss with 88% validation accuracy.

Below is the lineplots showing accuracy and loss:

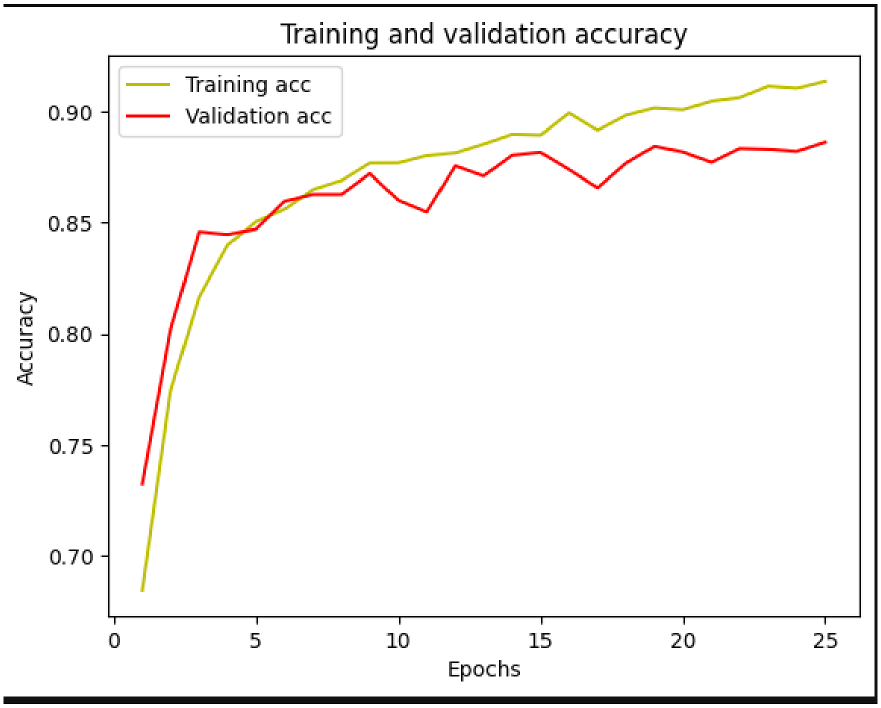


Figure 4.2: Training and Validation Accuracy for Gender

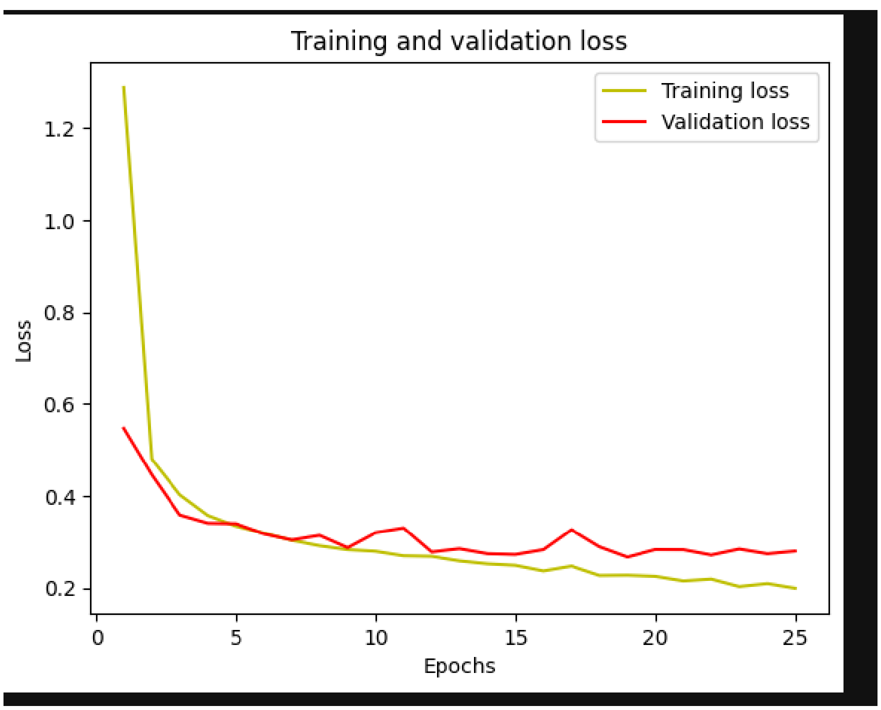


Figure 4.3: Training and Validation loss for Gender

Confusion Matrix

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm.
Below is the confusion matrix of the testing data yielded by the model for gender detection:

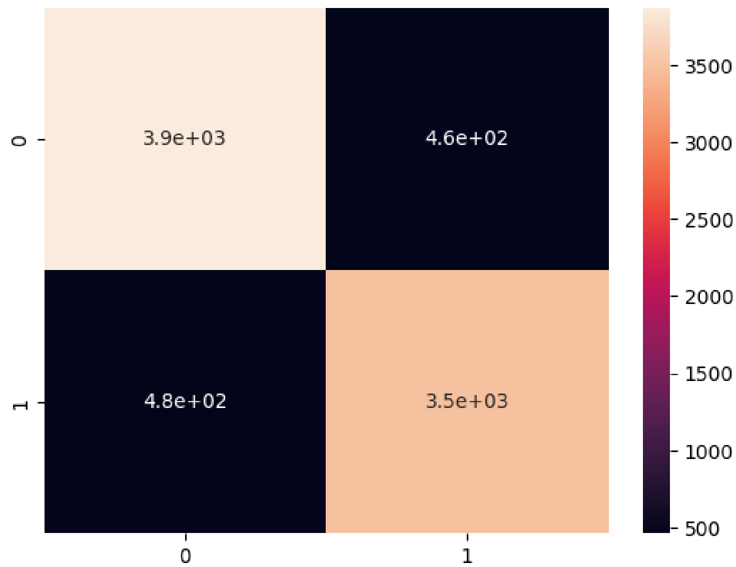
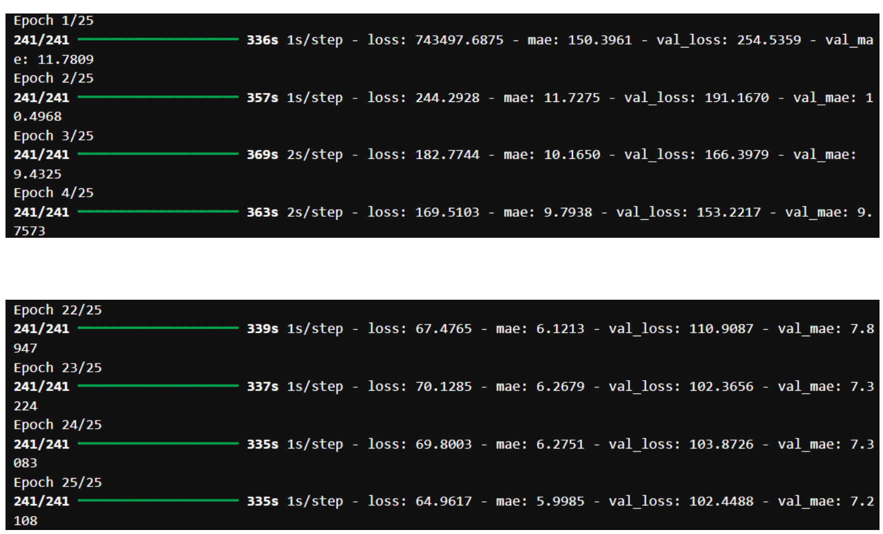


Figure 4.4: Confusion Matrix

4.1.2 Training Age Model

25 epochs were performed for model fitting that are shown below:



```
Epoch 1/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 336s 1s/step - loss: 743497.6875 - mae: 150.3961 - val_loss: 254.5359 - val_ma
e: 11.7809
Epoch 2/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 357s 1s/step - loss: 244.2928 - mae: 11.7275 - val_loss: 191.1670 - val_mae: 1
0.4968
Epoch 3/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 369s 2s/step - loss: 182.7744 - mae: 10.1650 - val_loss: 166.3979 - val_mae:
9.4325
Epoch 4/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 363s 2s/step - loss: 169.5103 - mae: 9.7938 - val_loss: 153.2217 - val_mae: 9.
7573
```

```
Epoch 22/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 339s 1s/step - loss: 67.4765 - mae: 6.1213 - val_loss: 110.9087 - val_mae: 7.8
947
Epoch 23/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 337s 1s/step - loss: 70.1285 - mae: 6.2679 - val_loss: 102.3656 - val_mae: 7.3
224
Epoch 24/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 335s 1s/step - loss: 69.8003 - mae: 6.2751 - val_loss: 103.8726 - val_mae: 7.3
083
Epoch 25/25
241/241 ━━━━━━━━━━━━━━━━━━━━ 335s 1s/step - loss: 64.9617 - mae: 5.9985 - val_loss: 102.4488 - val_mae: 7.2
108
```

Figure 4.5: Model Training process for Age Detection

- On first epoch, the loss was 743497.6875 with 150.396 MAE and 254.53 validation loss with 11.78 validation MAE.
- Coming to the 25th epoch, the loss was 64.96 with 5.998 MAE and 102.44 validation loss with 7.21 validation MAE.

Below is the lineplots showing accuracy and loss:

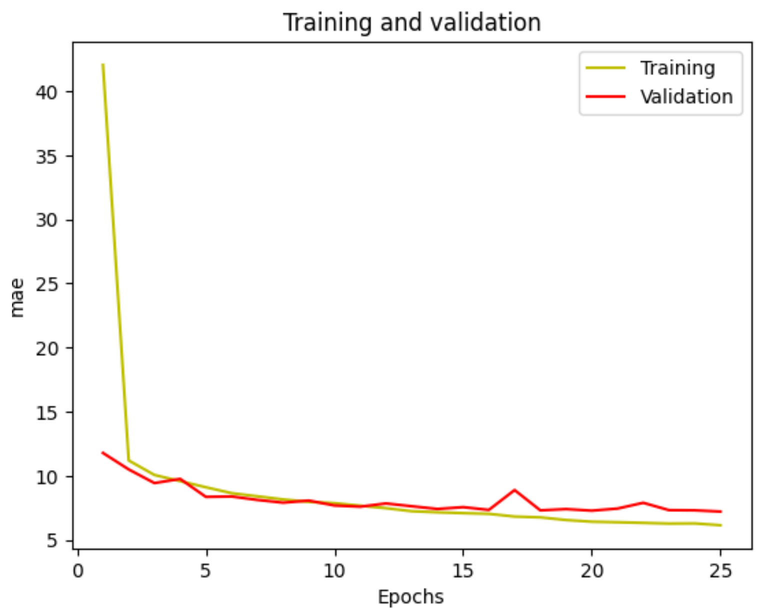


Figure 4.6: Mean Absolute Error for Age

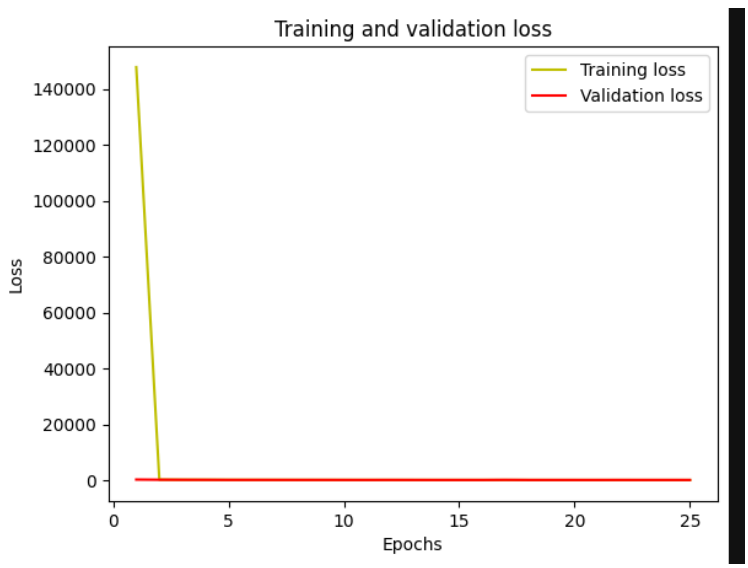


Figure 4.7: Training loss for age

4.1.3 Software Requirement

1. Python
2. vs-code and Jupiter notebook
3. React for frontend and Flask for Backend
4. OpenCV: We used OpenCV dependency for python . OpenCV is library where there are lots of image processing functions are available. This is very useful library for image processing. Even one can get expected outcome without writing a single code. The library is cross-platform and free for use under the open-source BSD license. Example of some supported functions are given bellow:

- Hough transforms: lines, segments, circles, and geometrical shapes detection
- Histograms: computing, equalization, and object localization with back projection algorithm
- Segmentation: thresholding,distance transform,foreground/background detection, watershed segmentation
- Filtering: linear and nonlinear filters, morphological operations
- Cascade detectors: detection of face, eye, car plates
- Interest points: detection and matching
- Video processing: optical flow, background subtraction, camshaft (object tracking)

5. Result And Discussion

The proposed system exhibits excellent performance by achieving a good classification of age and gender with reduced computation time and higher accuracy. The system receives the input picture in real-time via the camera. The source image is preprocessed to enhance the matching process's efficiency. Images are initially scaled at 200*200. The entry to the convolution network is 200*200 significantly. The convolutional layer applies a set of filters to the input image to extract important features and create a set of output feature maps. These feature maps contain information about the presence and location of specific features in the image. After each convolution is a MaxPooling layer which takes these output feature maps and reduces their spatial dimensionality by selecting the maximum value in each pooling window. This operation effectively down samples the feature maps, reducing their size while preserving the most important features. The activation function decides the value of pixels that help to build the model for the prediction of age and gender.
After building and training of the CNN models for age and gender prediction with UTK dataset, Haar-cascade classifier is used for the detection of faces and converted into a gray scale image for real time video by creating rectangle on the face. The gray scale image is reshaped into three channel for the input of the model. Gender and age model takes input from real time video and predict the age and gender of the face.

For the age detection model, the first epoch had a high loss of 743497.6875 and an MAE of 150.396 MAE, indicating poor performance. The validation loss was 254.53 with a validation MAE of 11.78, suggesting that the model was overfitting to the training set and performing poorly on new data. However, by the 25th epoch, the model had significantly improved, with a loss of 64.96 and an MAE of 5.998. The validation loss was 102.44 with a validation MAE of 7.21, indicating that the model was able to generalize well to new data and achieve a reasonable level of accuracy for age detection. These results indicate that the model was able to accurately estimate the age of the subjects in the dataset, with a mean absolute error (MAE) of approximately 10 years on the test set. On the other hand, the gender detection model had a better performance from the first epoch, with a loss of 4.5897 and an accuracy of 63%, and a validation loss of 0.5469 and a validation accuracy of 73%. By the 25th epoch, the model improved significantly, with a loss of 0.2024 and an accuracy of 91%, and a validation loss of 0.2799 and a validation accuracy of 88%. This indicates that the model was able to accurately detect gender from the data and perform well on new, unseen data, achieving an accuracy of 87% on the test set.
We had trained our model to 50 epochs too. As the training accuracy of the model got

increased but we don't get the significant increment in our validation accuracy. Infact we got some random increment and decrement in the accuracy in the increasing of the epochs.

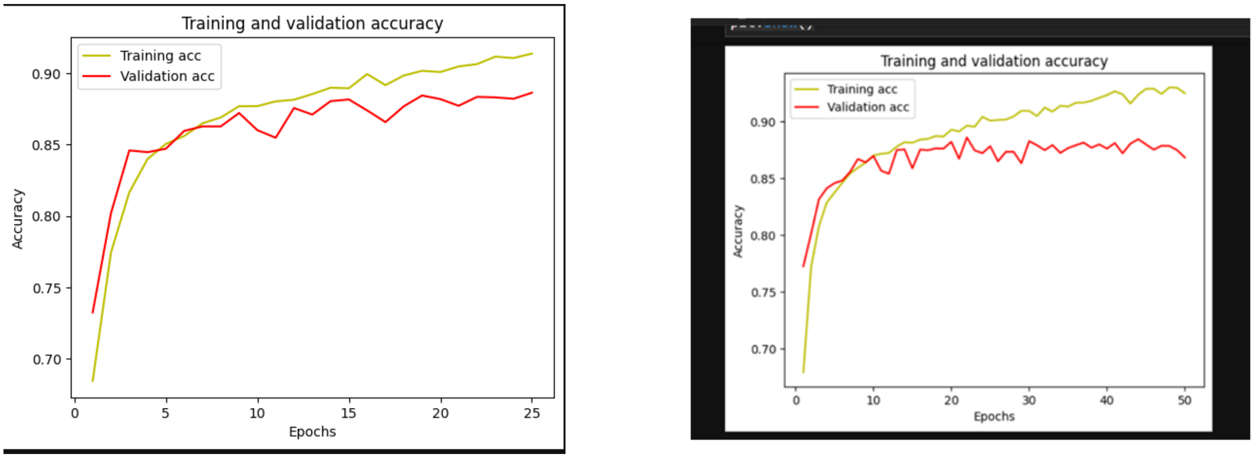


Figure 5.1: Comparision of the model accuracy of 25 and 50 epochs

The same case happened for mean absolute error (mae) for the model as mae doesn't gets significantly decreased during validation process on increasing the epochs.

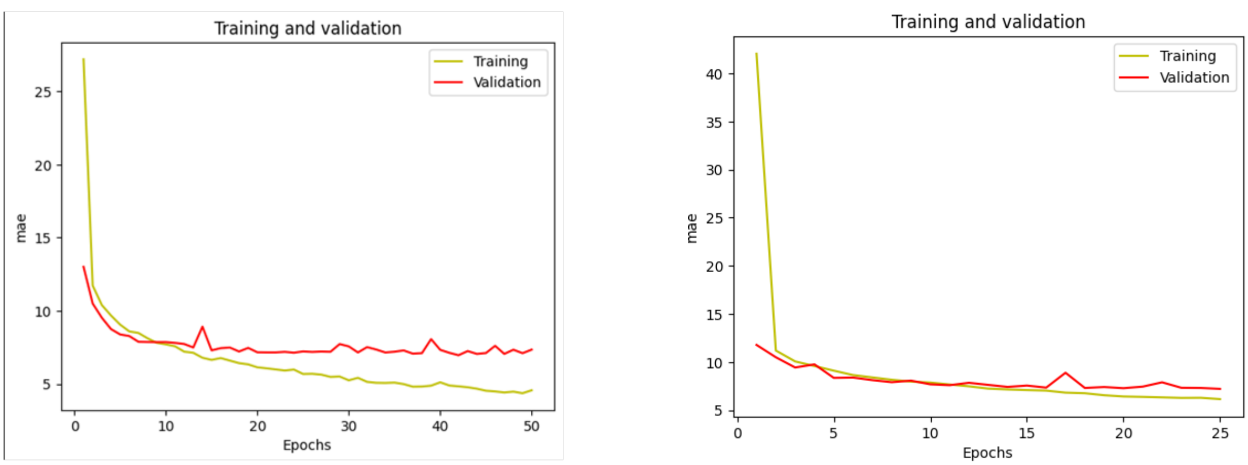


Figure 5.2: Comparision of the model mae of 25 and 50 epochs

The designed system is capable of detecting multiple faces from a frame.

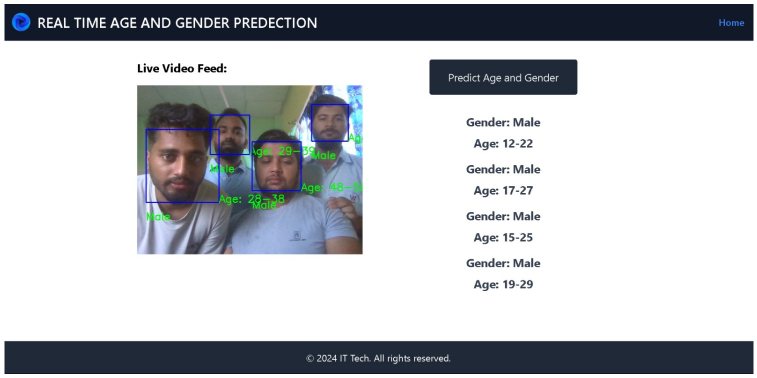


Figure 5.3: Multiple detection at a time

Overall, the results demonstrate that CNN models can be effective for gender and age detection tasks, with the ability to achieve high accuracy and generalize well to new data. However, further improvements could be made by using larger and more diverse datasets, fine-tuning hyperparameters, and incorporating additional techniques such as attention mechanisms or ensembling.

Analysis of Age and Gender Prediction on a Dataset of 50 Real Instances.

Age Prediction Accuracy

We conducted age prediction tests on a dataset of 50 real images, achieving an accuracy of 56.00% within a ±10-year range.

Actual Age \ Predicted Age	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
1-10	5	2	0	0	0	0	0	0
11-20	0	3	2	1	0	0	0	0
21-30	0	2	6	2	2	0	0	0
31-40	0	0	1	2	1	1	0	0
41-50	0	0	1	1	2	1	2	0
51-60	0	0	0	0	0	4	1	0
61-70	0	0	0	0	0	1	4	1
71-80	0	0	0	0	0	0	0	2

Figure 5.4: Confusion matrix of age

Gender Prediction Accuracy

We also tested gender prediction on the same dataset. The results were visualized using a confusion matrix, revealing an overall accuracy of 75.51%.

Actual Gender /Predicted Gender	Male	Female
Male	27	5
Female	7	10

Figure 5.5: Confusion matrix of gender

The results demonstrate the model's varying performance across age and gender predictions. While age predictions show room for improvement, especially for older individuals, gender predictions achieved a relatively high accuracy. These insights are valuable for further refining the model to enhance its accuracy and reliability in practical applications.

6. Limitation

Below are the limitations of the designed program:

- Input image resolution isn't good enough, which impacted the better result prediction.
- The brightness of the surrounding alters the results; input taken in dark surrounding has less accuracy compared to the input taken in bright surrounding.

7. Conclusion

We tackled the classification of age group and gender of unfiltered real-world face images. We used the UTKFace dataset and developed the model. Haar-Cascade algorithm was used to detect face's, binary cross-entropy for classification of gender and finally linear regression function for age detection. Training and testing accuracy was visualized using lineplots and confusion matrix (gender). The image preprocessing algorithm, handled some of the variability observed in typical unfiltered real-world faces, and this confirmed the model applicability for age group and gender classification in-the-wild.
We also see that after a certain number of epochs(around 20) the validation loss starts to fluctuate and eventually increases slightly, which might indicate overfitting. This suggests that while the model continues to improve on the training data, its performance on unseen validation data does not improve correspondingly. So we reduced the epochs from 50 to 25 where training is halted when the validation performance starts to degrade, to prevent overfitting and reduce the computation power of the system.
Model doesn't get exact ages of data as the facial expression can too change the feature of the data and can change accordingly. So putting them in the range of 10 years makes it appropriately by reducing the errors.
Hence, we conclude the report stating the objectives accomplished. Finally, we investigated the classification accuracy on UTKFace dataset for age and gender; the self-trained model achieved the state-of-the-art performance, in both age group and gender classification, significantly outperforming the existing model

8. Future Enhancement

Utilizing following methods could significantly improve the result of the system:

- Using a high speed processor and a better resolution camera with high focus efficiency/capability could improve the accuracy and also quickly identify/detect the face/s on the frame.

- Enhancing the dataset to train and also test the system with variety of images would improve the decision making ability of the system which could result in better performance.

- Training the model focusing on pre-processing, normalization, augmentation and multi-scale prediction could significantly reduce the effects of brightness of surrounding.

9. Timeline

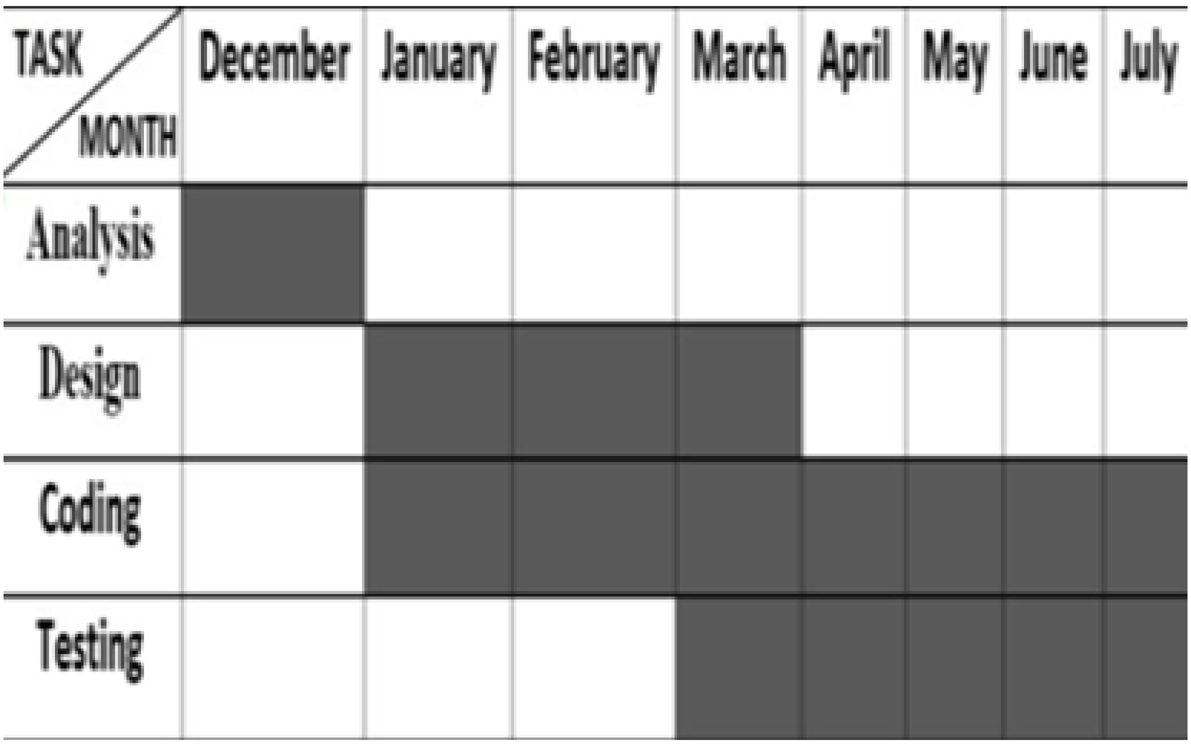


Figure 9.1: Timeline

References

[1] Ke Zhang, Ce Gao, Liru Guo, Miao Sun, Xingfang Yuan, Tony X Han, Zhenbing Zhao, and Baogang Li. Age group and gender estimation in the wild with deep ror architecture. *IEEE Access*, 5:22492–22503, 2017.

[2] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018.

[3] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on information forensics and security*, 9(12):2170–2179, 2014.

[4] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 34–42, 2015.

[5] Feng Gao and Haizhou Ai. Face age classification on consumer images with gabor feature and fuzzy lda method. In *Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings 3*, pages 132–141. Springer, 2009.

[6] Shuicheng Yan, Ming Liu, and Thomas S Huang. Extracting age information from local spatially flexible patches. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 737–740. IEEE, 2008.

[7] Yun Fu and Thomas S Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10(4):578–584, 2008.

[8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Young H Kwon and Niels da Vitoria Lobo. Age classification from facial images. *Computer vision and image understanding*, 74(1):1–21, 1999.

[11] Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):621–628, 2004.

[12] Asuman Gunay and Vasif V Nabiyev. Automatic age classification with lbp. In *2008 23rd international symposium on computer and information sciences*, pages 1–4. IEEE, 2008.

[13] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S Huang. Human age estimation using bio-inspired features. In *2009 IEEE conference on computer vision and pattern recognition*, pages 112–119. IEEE, 2009.

[14] Mohammad Ali Beheshti-Nia and Zahra Mousavi. A new classification method based on pairwise svm for facial age estimation. *Journal of Industrial and Systems Engineering*, 10(1):91–107, 2017.

[15] Ambra Demontis, Battista Biggio, Giorgio Fumera, and Fabio Roli. Super-sparse regression for fast age estimation from faces at test time. In *Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part II 18*, pages 551–562. Springer, 2015.

[16] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008.

[17] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010.

[18] Guodong Guo and Guowang Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *CVPR 2011*, pages 657–664. IEEE, 2011.

[19] MR Dileep and Ajit Danti. Human age and gender prediction based on neural networks and three sigma control limits. *Applied Artificial Intelligence*, 32(3):281–292, 2018.

[20] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[21] Dong Yi, Zhen Lei, and Stan Z Li. Age estimation by multi-scale convolutional network. In *Asian conference on computer vision*, pages 144–158. Springer, 2014.

[22] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D Barkana. Deep convolutional neural network for age estimation based on vgg-face model. *arXiv preprint arXiv:1709.01664*, 2017.

[23] Rajeev Ranjan, Swami Sankaranarayanan, Carlos D Castillo, and Rama Chellappa. An all-in-one convolutional neural network for face analysis. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pages 17–24. IEEE, 2017.

[24] Wei Liu, Lin Chen, and Yajun Chen. Age classification using convolutional neural networks with the multi-class focal loss. In *IOP conference series: materials science and engineering*, volume 428, page 012043. IOP Publishing, 2018.

Appendices

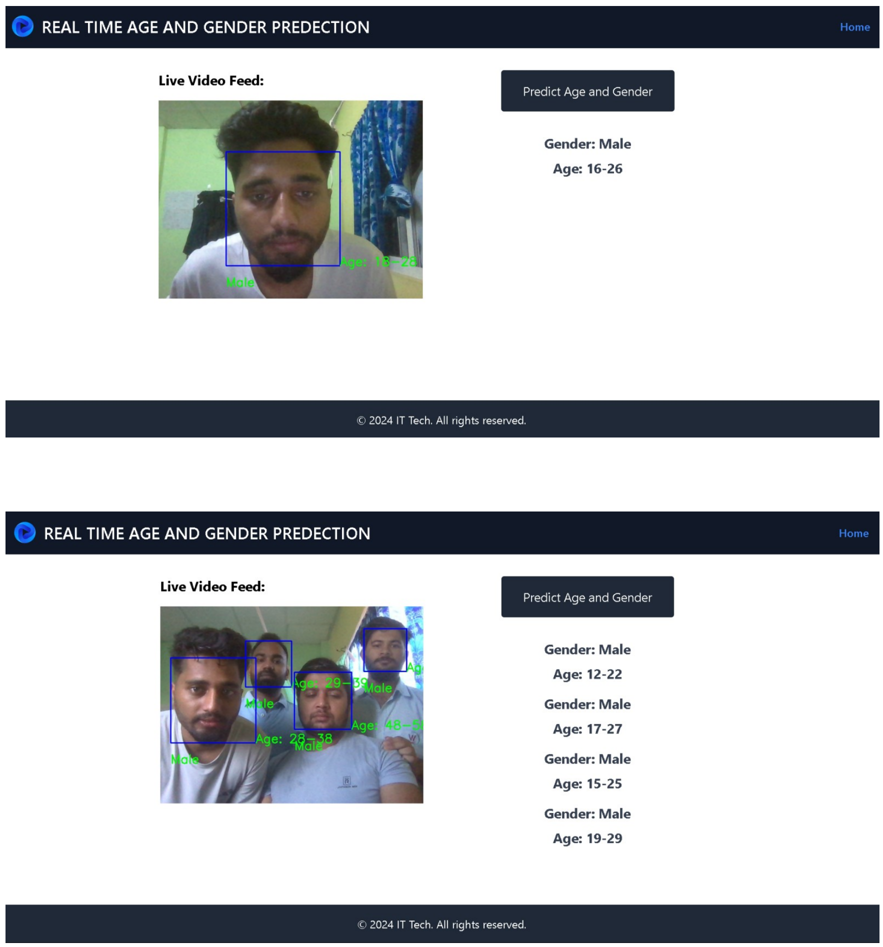


Figure 9.2: Predicted Outputs