Web application for Facial Expression, gender and age recognition

A Major Project Report Submitted in the
Partial Fulfillment of the Requirements
for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHONOLOGY

Submitted by

Alety Manish	**18881A1202**
Mylavarapu Soumya	**18881A1242**

SUPERVISOR
Mrs. N. Sujata Gupta
Assistant Professor

Department of Information Technology

VARDHAMAN COLLEGE OF ENGINEERING
(AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE, Accredited by NAAC with A++ Grade, ISO 9001:2015 Certified
Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

May, 2022

VARDHAMAN COLLEGE OF ENGINEERING
(AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE, Accredited by NAAC with A++ Grade, ISO 9001:2015 Certified
Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

Department of Information Technology

CERTIFICATE

This is to certify that the project titled **Web application for Facial Expression, gender and age recognition** is carried out by

Alety Manish	**18881A1202**
Mylavarapu Soumya	**18881A1242**

in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Information Technology** during the year 2021-22.

Signature of the Supervisor
Mrs. N. Sujata Gupta
Assistant Professor

Signature of the HOD
Dr. Muni Sekhar Velpurur
Professor and Head, IT

Project Viva-Voice held on ______________________

Examiner

Acknowledgement

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to **Mrs. N. Sujata Gupta**, Assistant Professor and Project Supervisor, Department of Information Technology, Vardhaman College of Engineering, for her able guidance and useful suggestions, which helped us in completing the project in time.

We wish to express our deep sense of gratitude to **Dr. M. Ramachandro**, Associate Professor and Project Co-ordinator, Department of Information Technology, Vardhaman College of Engineering, for his valuable suggesions and valuable inputs and assessment really helped us to shape this report to perfection.

We are particularly thankful to **Dr. Muni Sekhar Velpurur**, the Head of the Department, Department of Information Technology, his guidance, intense support and encouragement, which helped us to mould our project into a successful one.

We show gratitude to our honorable Principal **Dr. J.V.R. Ravindra**, for providing all facilities and support.

We avail this opportunity to express our deep sense of gratitude and heartful thanks to **Dr. Teegala Vijender Reddy**, Chairman and **Sri Teegala Upender Reddy**, Secretary of VCE, for providing a congenial atmosphere to complete this project successfully.

We also thank all the staff members of Information Technology department for their valuable support and generous advice. Finally thanks to all our friends and family members for their continuous support and enthusiastic help.

Alety Manish

Mylavarapu Soumya

Abstract

For a computer and human interaction, human facial recognition is crucial. Our goal is to anticipate the expression of a human face, gender, and age as quickly and accurately as possible in real time. Understanding human behaviour, detecting mental diseases, and creating synthetic human expressions are only few of the applications of automatic human facial recognition. Salespeople can employ age, gender, and emotional state prediction to help them better understand their consumers. Convolutional Neural Network one of the Deep learning techniques is utilized to design the model and predict emotion, age, and gender, using the Haar-Cascade frontal face algorithm to detect the face.This model can predict from video in real-time. The process of detecting face, pre-processing, feature extraction, and the prediction of expression, gender, and age is carried out in steps.

Keywords: Convolutional Neural Network; HaarCascade frontal face algorithm;

Table of Contents

List of Figures

Abbreviations

Abbreviation	Description
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
FER	Facial Expression Recognition
ReLU	Rectified Linear Unit
FC	Fully Connected
ANOVA	Analysis Of Variance
MANOVA	Multivariate Analysis Of Variance

CHAPTER 1

INTRODUCTION

1.1 Introduction

Our ambitions have risen since the arrival of modern technology, and they have no limitations. In today's world, there is a wide range of research going on in the field of digital imaging and processing image. The rate of progress has been exponential, and it continues to rise.

The facial expression of a person shows the person's mood, state of mind, thinking, and psychopathology, which is mainly the person-to-person communication. The seven major expressions that may be easily classified from the human face are anger, disgust, fear, happiness, neutral, sad, and surprise. The activation of several sets of face muscles expresses our facial emotions. These may look good but major signals from the human face show the mental state of the person. Applications for Age and gender classification include authentication through faces, security checks in public places, feedback systems, communication between humans and computers, amusement, and forensic labs.

The goal of this project is to develop an Automatic Facial Expression, age, and gender Recognition System through a web application that can recognize and categorize human facial photographs with a variety of expressions into seven different expression classes, 2 classes of gender, and 8 classes of age[1].
Labels for expressions

1. Neutral
2. Angry
3. Disgust
4. Fear

5. Happy
6. Sadness
7. Surprise

Figure 1.1: Seven basic facial expressions

Labels for gender

1. Male
2. Female

Labels for age

1. (0-2)
2. (4-6)
3. (8-13)
4. (15-20)
5. (25-32)
6. (38-43)
7. (48-53)
8. 60+

1.2 Problem definition

Generally, all the human emotions and thoughts can be seen on the human face. The same expression of the human can be expressed by different humans in different ways. Similarly, there are multiple humans of the same age and gender but there are some characteristics that help to differentiate between different age groups and gender. There are also similarities between them. Automatic recognition of the expression from the human face can be used in many areas like the interaction between computers and humans, feedback systems, video surveillance, and helps to know the mental condition of the human. Some of the challenges that facial expression systems might face include detecting the human face, extracting the facial features, and then the expression classification, gender classification, and age classification.

Convolution neural network which is one of the deep learning techniques is used to construct a facial expression, gender, and age recognition system. Anger, Disgust, Fear, Neutral Happy, Sad, and Surprise are the seven classes of expressions for humans that are used to classify the face of humans. The classifier for detecting the expression is trained and tested using the Kaggle data set FER2013 [2].

The expected gender is either ”male” or ”female,” and the expected age is either (0–2), (4–6), (8–12), (15–20), (25–32), (38–43), (48–53), or (60–100). (8 nodes in the final layer by using softmax activation function). The age and gender classifiers are trained on adience benchmark classification for age and gender. When we consider various features like lighting, makeup, and facial expressions, determining the precise age of a person instead of ranging it would be difficult. So, we have used a classification task instead of a regression task.

1.3 Objective of Project

The project's goal is to create a web application that uses a camera to capture a live human face and classify it into one of seven expressions, two ages, and eight age groups. We used the FER2013 dataset[3] for facial expressions and the Adience benchmark age and gender dataset[4] for age and gender recognition to build CNN for facial expressions, age, and gender classification.

A convolutional neural network is one of the deep learning techniques that follow a feed-forward approach and is similar to the human brain and acts like the human brain in actions like thinking, reading images, and classifying images. Our motto is to design a convolutional model that reads the human face and categorizes based on the features it extracted from images and knowledge gained from it. The process of convolutional also approximates the result obtained from each neuron. Convolutional networks have multiple layers in their model, each layer uses the minimum time to pre-process. Each layer extracts many features to improve its learning ability.

1.4 Limitations of Project

For the model's facial expression, all the expressions are recognized correctly but the dataset considered has very less images for the label disgust. Since they are very less, the training for that label is not sufficient enough to predict the label correctly.

When we consider various features like lighting, makeup, and facial expressions, determining the precise age of a person instead of ranging it would be difficult. Improving the brightness and variations in the RGB values of the training images can improve the age and gender prediction in web-camera-based testing. Each photo in training and test data has two labels, i.e. gender label and age group label. Some photos in the Adience data set do not have gender labels or age labels.

1.5 Modules

There are totally four modules :

- Expression-classification-cnn-using-keras
- Gender recognition
- Age recognition
- Flask

Expression classification

This module is about training the model with various convolutional layers, and then dense layers, finally applying the softmax activation function whose output is the probability of seven expressions. The model is stored in a .h5 file which can be directly used later. Later it plots the loss and accuracy of the model.

Gender recognition

This module is about training the model with various convolutional layers, and then dense layers, finally applying relu and softmax function whose output is the probability of both genders(male and female). Later it plots the loss and accuracy of the model.

Age recognition

This module is about training the model with various convolutional layers, and then dense layers, finally applying relu and softmax function whose output is the probability of eight classes of age. Later it plots the loss and accuracy of the model.

Flask

This module uses the .h5 file generated in the previous module which has the trained model. It used OpenCV to detect faces through a webcam and classify the face into one of the expressions, age, and gender.

Flask is used here to integrate the trained models of deep learning with the website. Proper routing is maintained to move from one page of the website to other pages.

CHAPTER 2

LITERATURE SURVEY

2.1 Explain about existing system

Recognize Facial Expression using Binary patterns and cognition

This research paper proposed an approach to recognizing expressions from faces using binary patterns and cognition. They observed that the eyes and mouth are the main parts to classify an expression. The procedure begins by extracting facial contours using the LBP operator. They used a 3D model to divide the face into six sub-regions. They used the mapped LBP approach to recognize the expression from the sub-parts they divided earlier. They have used a support-vector machine and softmax activation function for two types of expression classification models namely basic expression model and circumplex expression model. Finally, they have compared the facial expression dataset Cohn-Kanade (CK +) with the test dataset in which they have considered ten members to compare. They observed that their model will remove the image's problematic elements. The expression model outperforms the traditional emotional model by a wide margin.[5]

Multicultural Facial Expressions Analysis using Artificial Neural Network

This project is to determine expression by building an artificial neural network based on different cultures. Variations in the appearance of the face, the structure of the face, and facial emotion representation due to cultural differences are the main problems with the facial emotion detection system. Because of these differences, multicultural face expression analysis is required. Several computational strategies are presented in this paper to address these changes and achieve excellent expression recognition accuracy.[6]

For intercultural facial expression analysis, they presented an artificial neural network-based ensemble classifier. By merging facial images from the Japanese female facial expression database, the Taiwanese facial expression image database, and the Radboud faces database, a multi-culture facial expression dataset is constructed. Members of the multicultural dataset are Japanese, Taiwanese, Caucasians, and Moroccans, who represent four ethnic groupings. Local binary patterns, uniform local binary patterns, and principal component analysis are utilized to express facial features

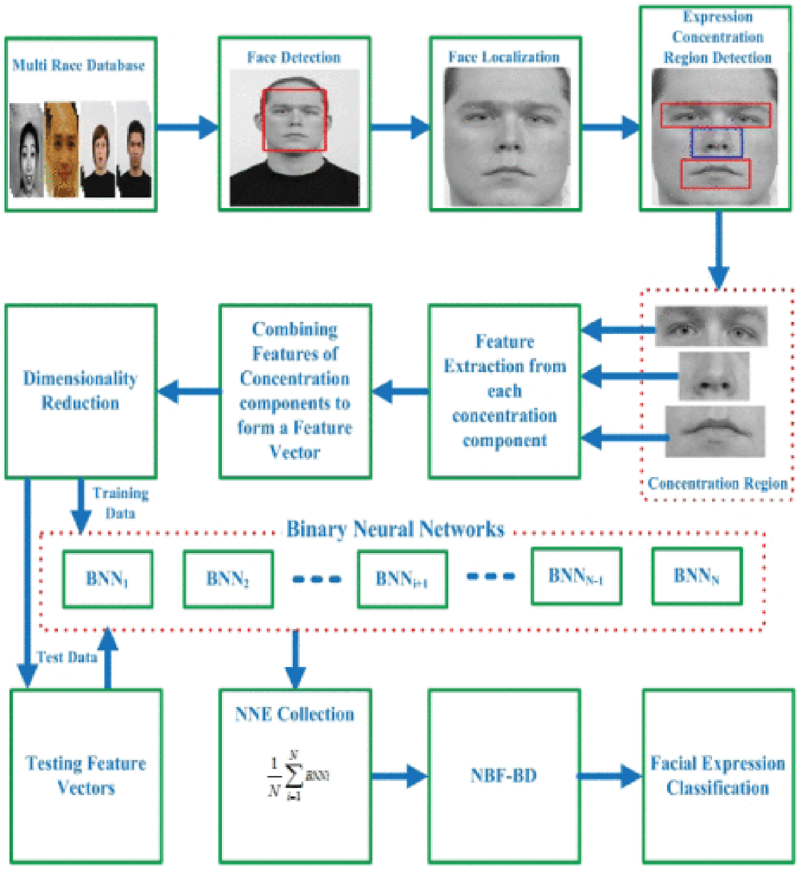

Figure 2.1: Framework for Multicultural Facial Expressions Analysis

Gender and Age recognition using speech

They have developed a model that recognizes the voice and classifies into age and gender. There are many factors that affect the process of automatic speech recognition like the speaker's weight, height, and speech based on a person's mood. There is also a requirement for a very large database that consists of a large number of speakers of different ages and genders. Since there will also be a problem of having noise when trying to record audio of the speaker, high-quality microphones and filters are also required. The result varies when different speakers are used to recording the same statement from the same speaker.

Automatic gender and age recognition can be done in a variety of ways. Cepstral characteristics, such as Mel Frequency Cepstral Coefficients, are an example (MFCC). For the purpose of age recognition. With recorded data, MFCC is known for delivering poor gender and age categorization results.To avoid this issue, the MFCC features are improved by examining the parameters that influence the feature extraction process. MFCC has been employed in a variety of speech applications, including voice recognition and language recognition, and another acoustic characteristic that can be derived is format frequency[7].

Emotion Recognition using body movements

This project is to detect the emotions from the body movements. They have used these body movement characteristics and developed a two-feature selection framework. They have considered only five basic features namely anger, happiness, sadness, fear, and neutrality. The first layer is the combination of Multivariate Analysis of Variance (MANOVA) and Analysis of Variance (ANOVA) to remove unnecessary features for emotion detection. In the second layer, they have used a binary chromosome-based genetic algorithm to pick the relevant features that improve in giving the correct label of emotion. Some of the movements they have considered are walking, sitting, and action-independent instances.

Based on experiments of various body movements, this model is outperformed in terms of emotion recognition rate.The acurracy of the action walking is 90 percent, 96 percent for sitting and other action-independent scenarios accuracy is 86.66[8].

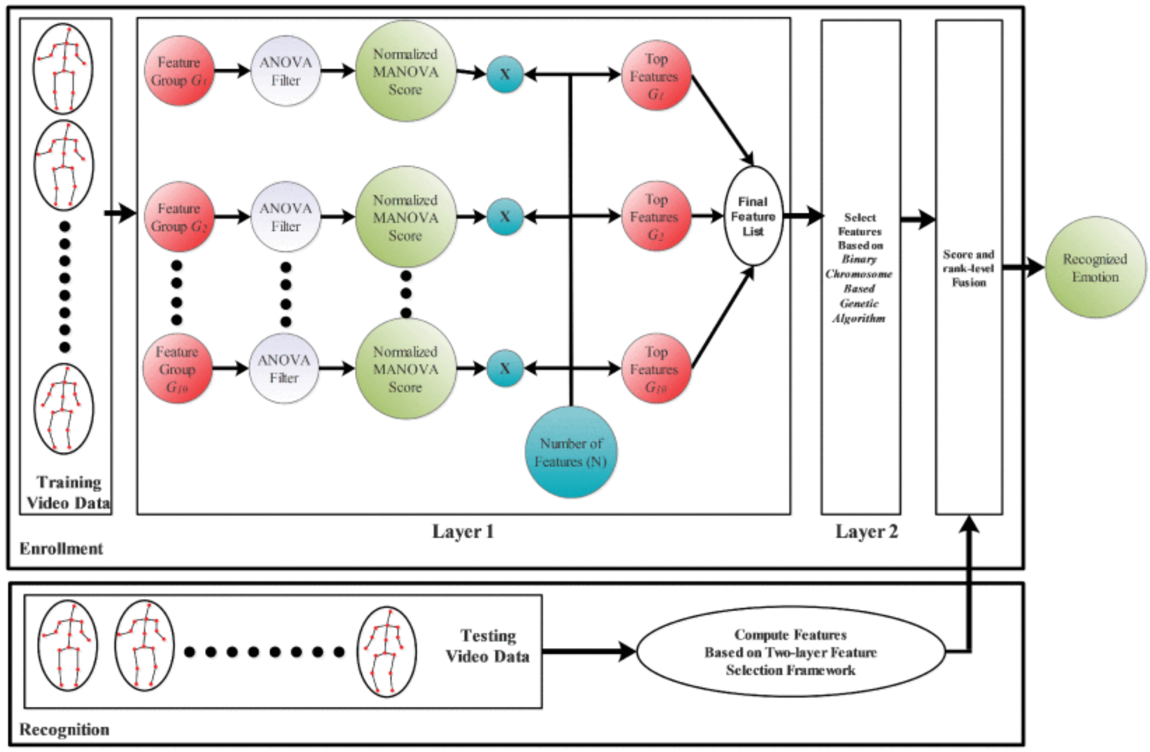

Figure 2.2: Proposed method for emotion recognition using body movements

2.2 Limitations of Existing System

LBP can achieve impressive accuracy in pattern recognition fields with a strong texture discrimination capability. However, in the real-world application, the basic LBP has two main limitations: sensitivity to local illumination variations and huge feature dimensions. Furthermore, it is unable to adequately characterize the texture of face muscles, wrinkles, and other local deformations, and so does not always perform well in FER.

There are challenges that arise in the process of automatic age estimation based on speech. Firstly, the factors like a person's mood, speaker's height, and weight affect the voice. Secondly, the database should be very huge and consists of the different speeches of different age groups. Lastly, the speech that we hear daily will have a lot of disturbances and noise. To avoid that kind of noise getting recorded, speakers of high quality are required which are very costly. When different recording methods are used to record the same statement from the same speaker, the results vary.

From the model that uses multicultural databases to train their model, the human faces of different cultures might not get the labels correctly. This shows that accuracy for the other cultures humans is less compared to the cultures of the human that the model is trained on.

The model which is built based on the body movements of the human also has some limitations. The body movements of a person are not restricted only to a finite number. The body movements of all the humans are infinite and if the model encounters any movement that it couldn't compare with its learning, then its accuracy in predicting those labels will be less.

2.3 Proposed Method

Facial Expression, age, and gender Recognition takes in input as an image through a web camera, detects the face, and then classifies the face into one of the seven basic expressions, one of the two genders, one of the eight class labels of age.

For detecting the face, we have used the haar cascade frontal face algorithm and used convolutional neural networks to train the model.

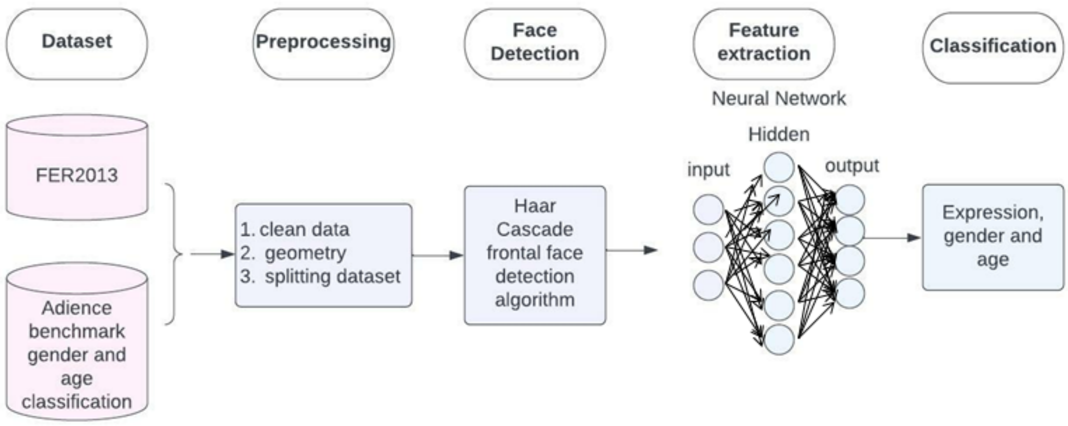

Figure 2.3: Proposed method for expression, gender and age reccognition

CHAPTER 3

ANALYSIS

3.1 Software Requirement Specification

3.1.1 User requirement

The user requirements specify what the user require from the system.

1. The system should be able to recognize the face and classify it into one of the expressions, age and gender.
2. Accuracy of the model should be high
3. Model must provide the quality of service to user.

3.1.2 Software requirement

- Jupyter notebook
- Python 3

Libraries required are

- tensorflow
- keras
- matplotlib
- pandas
- flask

3.1.3 Hardware requirement

- Processor : Intel CORE i3 processor with minimum 2.9 GHz speed.
- RAM : Minimum 4 GB.

- Hard Disk : Minimum 500 GB

3.2 Algorithms and Flowcharts

Algorithm:

1. Collecting a data set of photos is the first step. (We're utilising the FER2013 database, which contains 35887 pre-cropped, 48-by-48-pixel gray-scale photos of faces, each classified with one of the seven emotion classes: anger, contempt, fear, happiness, sadness, surprise, and neutral. Other dataset we are using is Adience benchmark age and gender classification which has images of different ages and different age groups)
2. Image pre-processing. (Pre-processing includes, setting the geometry of the training images, separating the dataset for testing and training, removing the noisy data, etc.)
3. Identifying a face in each image.(OpenCv library to detect the face)
4. The gray-scale photographs of the clipped face are created.
5. The pipeline ensures that each image may be fed as a (1, 48, 48)numpy array into the input layer for expression and (1, 56, 56) for age and gender.
6. The Convolution2D layer receives the numpy array. Feature maps are generated by convolution.
7. MaxPooling2D is a pooling approach that uses (2, 2) windows across the feature map to maintain only the highest pixel value.
8. During training, the pixel values were subjected to forward and backward propagation by neural networks.
9. For each emotion class, the Softmax function appears as a probability. The model can display the probabilistic composition of the expresions, gender, and age in the face in great detail.

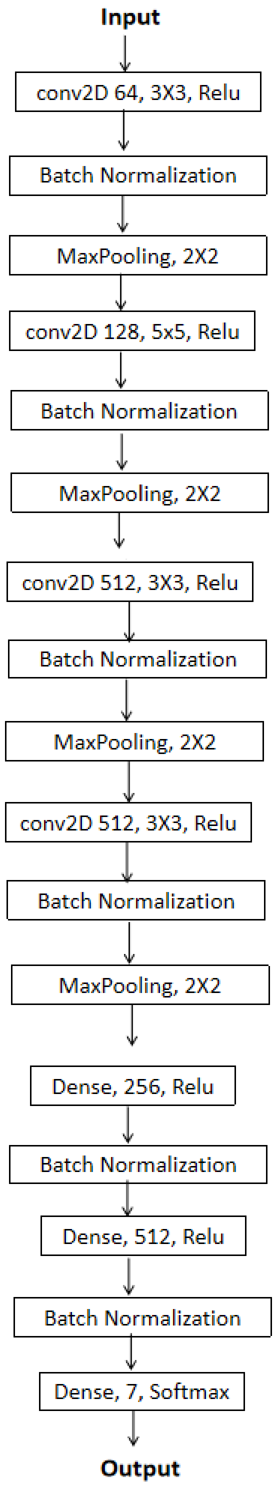

Figure 3.1: Flow chart

CHAPTER 4

DESIGN

4.1 Introduction

The deep learning model is built which will give the images' emotion, gender, and age state. The training data set is used to build the model and learn from this data to set the parameters. The testing data is used to evaluate the performance of the model. The testing data is completely unobserved data. The output from the training data is present in the model and the data learned from the training data is used to predict the label for testing data. If the labels for the testing data predict correctly, then we can get to the conclusion that the model got trained very well since it gives the correct label for the data that it didn't come across before. To send the data for the training, first, it must be pre-processed.

Pre-processing

Pre-processing is the first step to be followed to train and build the model. Performing image pre-processing never increases the information of the image but it decreases the information making it easy for the model without affecting the training. The main aim of pre-processing is to reduce the information of images that causes disturbances during training and improve those features that are helpful for training and future process.

There are 4 different types of Image Pre-Processing techniques and they are listed below.

1.Pixel brightness transformations/ Brightness corrections

2.Geometric Transformations

3.Image Filtering and Segmentation

4.Fourier transform and Image

The brightness of the pixel is changed in the brightness transformation. This transformation depends on the value of the pixel. In this technique, the output of the pixel value is based on the input value of the pixel. Operations that make changes in brightness include color correction, adjustments of brightness, etc. Major changes in brightness improve image processing and help in the field of computer vision. Some of the areas where pre-processing steps includes medical image processing, recognition of speech, and applications that deal with images and videos.

There are two types of Brightness transformations and they are below.
1.Correction of brightness
2.Gray scale transformation

In our project, the image brightness is transformed into a gray scale. Facial expression recognition is a process performed by humans or computers, which consists of:

1. Locating faces from the image or video is called face detection.

2. Extract the features from the face that is recognized in the previous step. Detect the shape of the components of the face like eyes, eyebrows, mouth, nose, skin texture, etc. This is generally called facial feature extraction.

3. To observe and analyze the movements and changes of the extracted features from the face. Some of the facial movements and changes include skin color, eye length, mouse length, the distance between eyes, the distance between eyebrows, alignments of eyes, node, mouth, etc. Use this information to classify into facial expression, gender, and age. For example, the wide mouth is a smile, and come to the conclusion that the expression is happy.

A computer technology called face registration is widely used in many applications like authentication, surveillance, etc to find the human face from the image that is captured digitally. In this step, faces are identified from the image by techniques for ”face detection” or ”face localization”. After the face is detected, then that image is normalized with any of the template images called “face registration”. [9]

Facial Feature Extraction

Feature extraction from the face is defined as the process to locate the points, regions, and landmarks from the image. A numerical feature vector gets generated in this feature extraction from an image. Some of the features include

a. nose

b. lips

c. Eyebrows

d. eyes

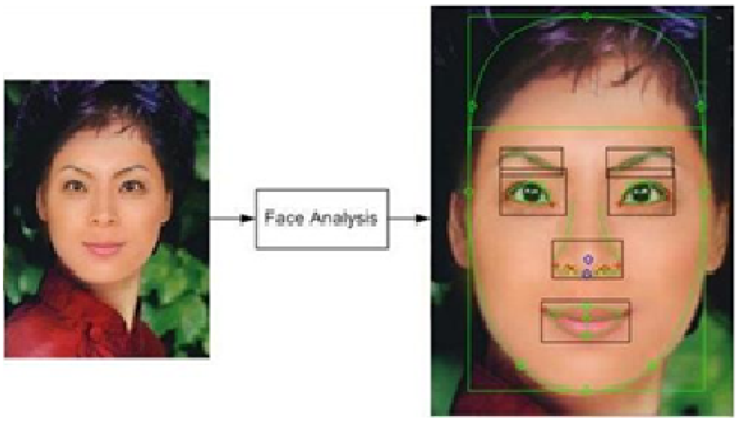

Figure 4.1: Facial feature extraction

Expression, gender and age Classification In the third step, of classification, the algorithm attempts to classify the given faces portraying one of the seven basic expressions, two genders, eight ages.

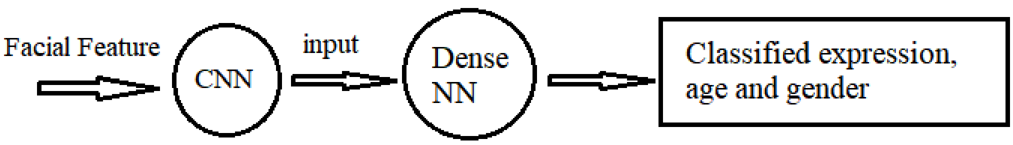

Figure 4.2: Classification

4.2 Data Flow diagram

The data flow diagram depicts our model. The first stage, pre-processing takes input as testing and training data set and processes the image like resizing, geometry setting, brightness transformation, splitting of training and testing dataset etc. That output is given to a face detection algorithm which is Haar Cascade frontal face detection algorithm which detects the face and then extracts the features from the face detected. After extraction of the features, expression, gender and age are classified based on the trained model. After classifying, outputs the label which is having the highest probability.

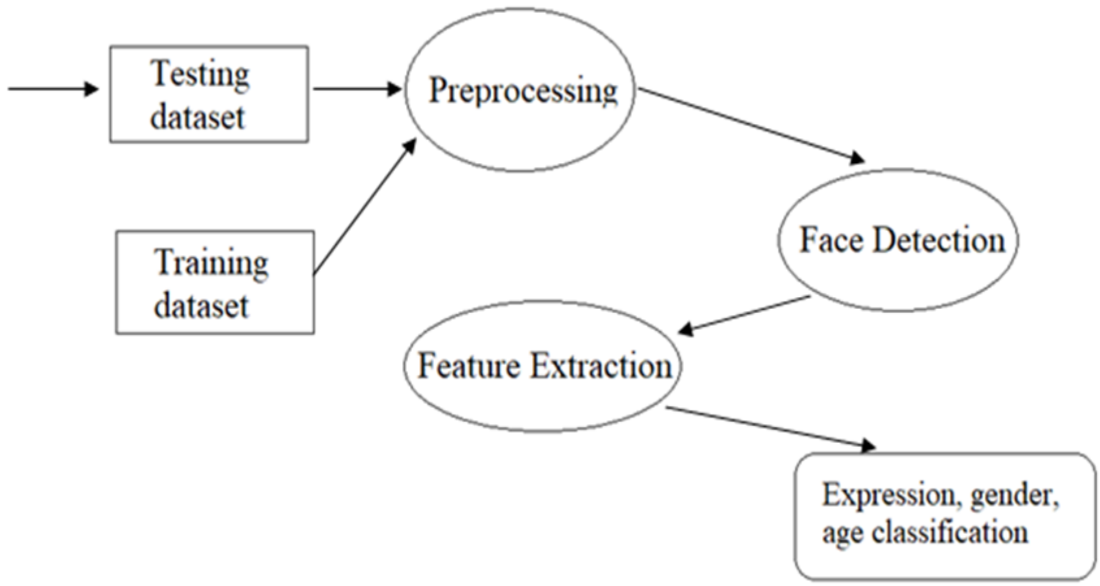

Figure 4.3: Data flow diagram for building model

For working of application, first when the flask application is run, then an ip address gets generated and copy this address and paste in browser to open the website designed. The website has the button,

clicking on the button navigates to the other route that lively captures the images from the web camera of the device. Internally the models that learned works and gives the labels for expression, age and gender for face detected.

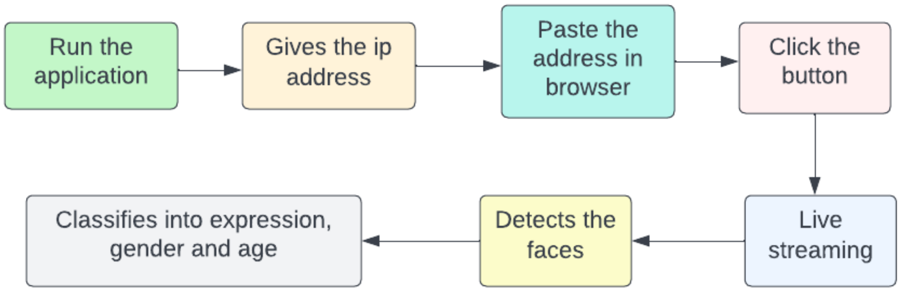

Figure 4.4: Data flow diagram for working of application

CHAPTER 5

IMPLEMENTATION

5.1 Introduction

People have ability to know the other's expression, gender and age. Similarly, the model must be built to give the gender, age and expression label. To implement the model that recognize the face and classify into expression, gender and age, neural networks are used. We have used Convolutional Neural Network to build the model that learns the features and gives the expected labels. A Convolutional Neural Network takes the input as an image which consider the pixel values from the image and assignes weights to them , then extracts the features from it and differentiate them. Next, the Haar Cascade algorithm is used to detect the face or identify the face from a real-time camera or image.

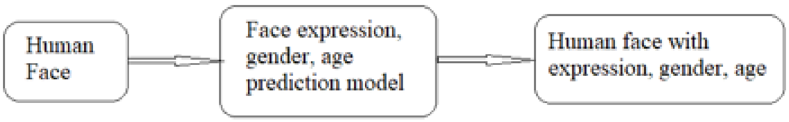

Figure 5.1: Implementation of Facial Expression recognition

5.2 Explanation Of Key Functions

5.2.1 Convolutional Neural Network

The Convolutional Neural Network (CNN) takes the image as input. This method assigns weights to objects of the image, which helps to distinguish one object from the other based on the differences it observed. When compared to other approaches, pre-processing for this approach is less[10, 11, 12]. After the training, the model extracts the features from the images by applying the filters and uses those features for its learning.

Convolution architecture has one input layer, one or more hidden layers, and one output layer. In this way, a bunch of layers in a convolutional neural network turn into output giving the probabilities of each class label. For each layer, there would be CONV, applying activation functions like RELU, Softmax, applying to pool like max-pooling, mean pooling. There might be or might not be parameters for each layer. Each layer takes the input from the previous layer's output. Each layer performs a differential function to give the output to the next layer[13].

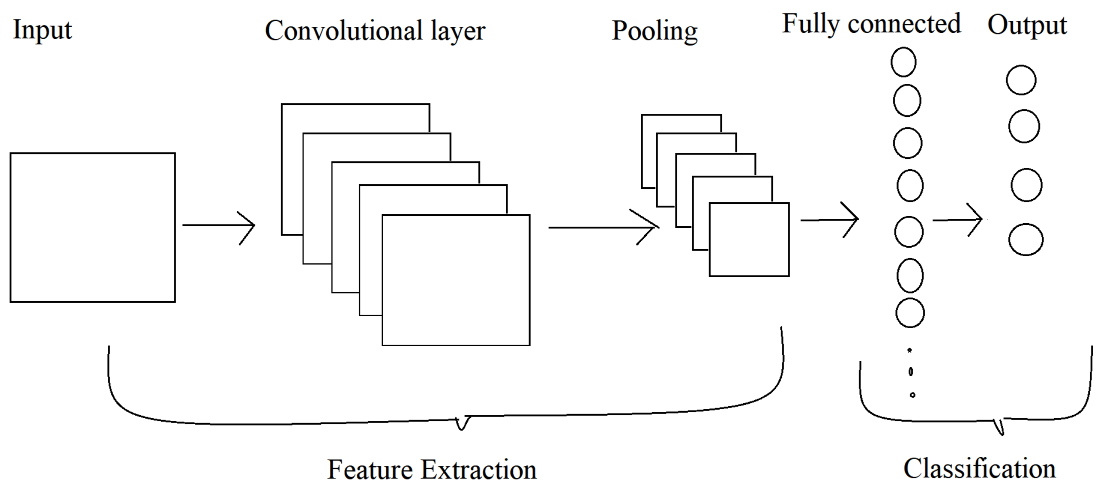

Figure 5.2: Convolutional Neural Networks

5.2.2 Haar Cascade Classifier

Now, Facial recognition is present in every place namely security cameras, and sensors on iPhone X. There are many different human faces where there are many differences between human to human and may also have similarities. How does facial recognition work to classify all the faces?

Haar Cascade classifiers are the classifiers that detect the face of humans in real-time and are able to differentiate between humans and non-humans. Haar classifier is a machine learning algorithm that accepts input as an image or video and recognizes objects. Haar Cascade's model got trained on many positive images that consist of objects that the classifier wants to identify and also have negative images that consist of objects that the classifiers don't want to recognize. Haar features are used to find how a given point is part of an object. To get a strong prediction, a group of weak learners is used with boosting algorithms. These algorithms are executed on multiple sections of an input image or video using cascade classifiers.[14, 15]Open CV can be used to implement Haar Cascade model. [9]

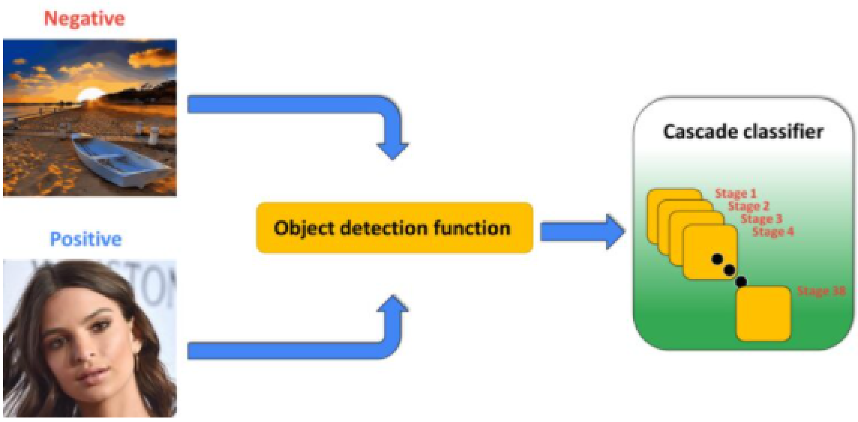

Figure 5.3: Haar Cascade Classifier

5.3 Technology

5.3.1 Convolutional Neural Networks

The neurons of the convolution layers have biases and weights. Each neuron of the layer takes input from each neuron of the previous layer and performs the operation dot product, then executes a non-linear function.

Convolutional Layers

The convolutional layer extracts the features from each layer and each operation that performs and then uses these features to learn the feature representations of the images that are given as input. In each feature map, each neuron is coupled with every neuron present in the previous layer by the set of the trainable weights, which is called a filter bank. To compute a new feature map, the weights from the layer are taken and these results are used in the activation functions like relu, softmax, tanh, etc. The feature maps of the same layer of convolutional have different weights, that extract multiple features at each location[13].

Pooling layers

The objective of the pooling layers is to minimize the shape of the feature maps. Average pooling is a method that calculates the average of all the input values of the previous output of the image, this is initially used. Max pooling is to find the maximum value of the considered dimension of the input and pass these values to the next layer. Using a pool of size 2 for max-pooling reduces the dimension by two

.

Fully Connected layers

All the convolutional layers, and pooling layers are stacked one after the other. The fully connected layer multiplies the weights of the input and then adds the bias to that value. The convolutional neural network can have any number of fully connected layers. The neurons of the fully connected layer connect with each neuron of its previous layer. The softmax activation function is the most commonly used function for classification algorithms. The study proves that to improve the accuracy, the softmax function is used with a support vector machine.

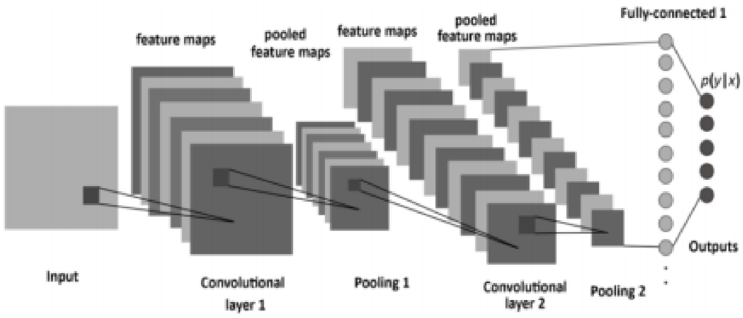

Figure 5.4: Implementation of CNN

5.3.2 Face Detection using Haar Cascade Classifier

Haar Cascade algorithm helps in detecting the face through live camera or an image.To train the classifier, the method requires a large number of images of faces(positive images) and images without faces (negative images). After that, we must extract features from it. This algorithm is implemented in four stages :

1. Calculating Haar Features

2. Creating Integral Images

3. Using Adaboost

4. Implementing Cascading Classifiers

Calculating Haar Features

Haar is a feature that commonly focuses on the edges of the images. All the human faces have similarities in the regions like eyes are darker than the region above the eyes, compared to the eyes, the nose region is darker. The placement of all these regions is the same for all humans. The model matches these kinds of features from the image to detect the human face.

[16].

A Haar feature is a set of calculations performed on consecutive rectangular sections of a detection window at a specified position. The calculation is adding up the pixel intensities in each region and then subtracting the total.These features can be difficult to determine for a large image. Integral pictures are useful in this situation since they reduce the number of operations required. **Creating Integral**

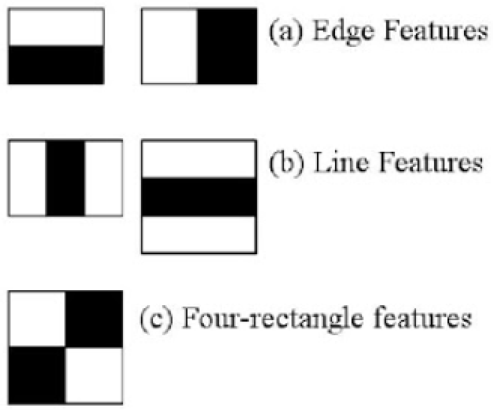

Figure 5.5: Haar Features

Images

The calculation of these Haar features is sped considerably by using integral pictures. Rather than computing at each pixel, it divides the screen into sub-rectangles and creates array of references for each of them. The Haar features are then computed using them[16].

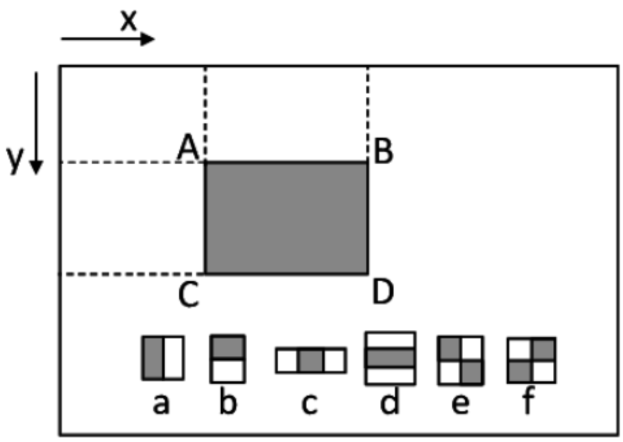

Figure 5.6: Illustration of how an internal images works

Using Adaboost

Adaboost basically selects the most useful features and trains the classifiers to use them. It creates a "strong classifier" by combining "weak classifiers" that the algorithm may use to detect items. Using cascading classifiers, the final phase merges these weak learners into strong learners.

Implementing Cascading Classifiers

The cascade classifiers have a number of stages in a sequence that

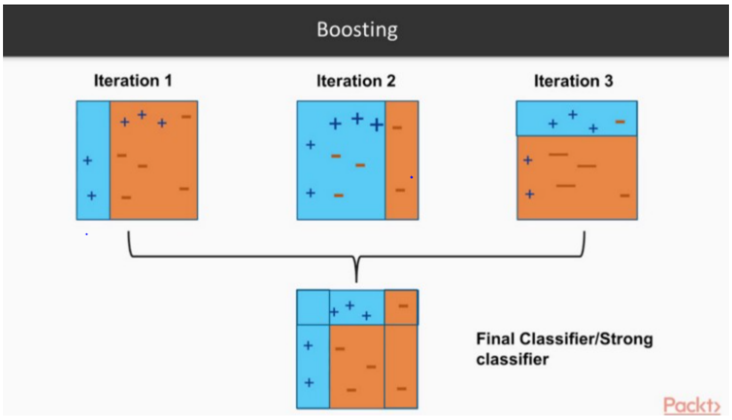

Figure 5.7: Adaboost algorithm implementation

includes a group of weak learners. Weak learners are those who couldn't learn properly and couldn't predict correctly. The weak learners are taught via boosting, which produces a highly accurate classifier based on the average prediction of all weak learners.[16].

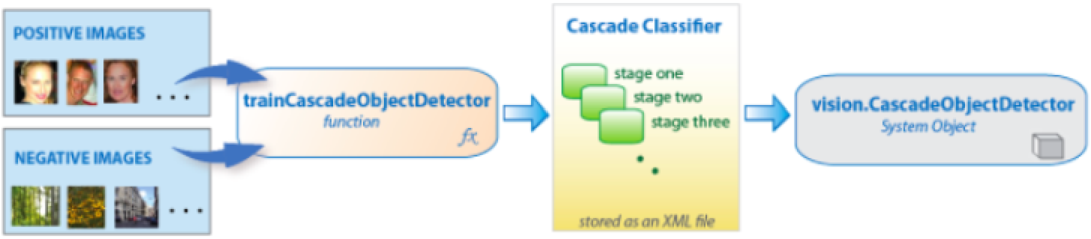

Figure 5.8: Cascade Classifiers

The classifier then selects whether to mark an object as found (positive) or to move on to the next region based on this prediction (negative). Because the bulk of the windows does not contain anything of interest, stages are designed to reject negative samples as quickly as feasible. Because classifying an object as a non-object can severely impair your object identification method, it's critical to maximize a low false-negative rate.

5.4 Method of Implementation

5.4.1 The Database

We used a data set from a Kaggle Facial Expression Recognition Challenge a few years ago to train the model (FER2013). It consists of 35887, 48x48 grayscale images of faces that have been pre-cropped and classified with one of the seven emotion classes: anger, disgust, fear, happiness, sorrow, surprise, and neutral. There are multiple images for each expression. One picture from each of the expressions is shown below.

Figure 5.9: Pictures from dataset with seven expressions

We used Adience benchmark age and gender classification dataset for age and gender recognition which consists of images of various genders and age classes. There are multiple folders in where there are multiple of images of human with different age and genders. The labels for those images in the folder are present in other files.

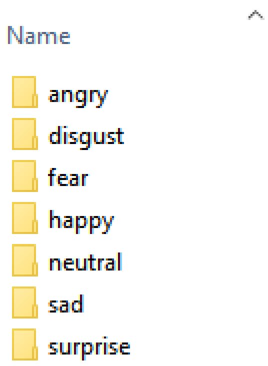

Figure 5.10: Folders with images of seven expressions

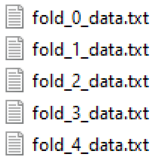

Figure 5.11: Files with labels of age and gender

user_id	original_image	face_id	age	gender
30601258@N03	10399646885_67c7d20df9_o.jpg	1	(25, 32)	f
30601258@N03	10424815813_e94629b1ec_o.jpg	2	(25, 32)	m
30601258@N03	10437979845_5985be4b26_o.jpg	1	(25, 32)	f
30601258@N03	10437979845_5985be4b26_o.jpg	3	(25, 32)	m
30601258@N03	11816644924_075c3d8d59_o.jpg	2	(25, 32)	m
30601258@N03	11562582716_dbc2eb8002_o.jpg	1	(25, 32)	f
30601258@N03	10424595844_1009c687e4_o.jpg	4	(38, 43)	f

Figure 5.12: Labels of age and gender with the image file name

5.4.2 Model

After the data is picked from the dataset, the images are sent to pre-processing. In pre-processing, images are converted such that they all are of the same geometry before they are sent as an input model. Noisy data is also removed from the dataset and splitting of data for training and validation also takes place. The face from the image is captured using the Haar-Cascade frontal face algorithm. After this phase, a convolutional neural network is generated which is having an input layer, one or more hidden layers, and an output layer.

Deep learning is a common computer vision approach. Convo-

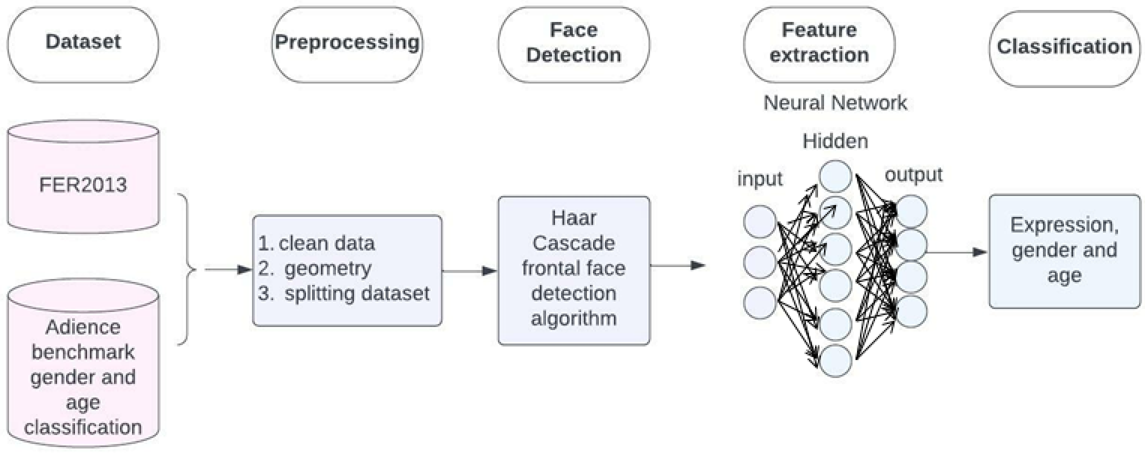

Figure 5.13: Implementation Proposed framework

lutional Neural Network (CNN) layers were chosen as the building blocks for our model architecture. When processing pictures, CNN's are known to mimic how the human brain functions.

A convolutional neural network's usual architecture includes an input layer, convolutional layers, some dense layers (also known as fully-connected layers), and an output layer. These are stacked layers that are organized linearly. The model is constructed as Sequential() in Keras, and further layers are added to build the architecture.

Input layer

Because the input layer's dimensions are pre-determined and fixed, the image must be pre-processed before being fed into it. For face detection in the image, we utilised Open CV, a computer vision package. OpenCV's haar-cascade frontalface default.xml file contains pre-trained filters and employs Adaboost to locate and clip the face quickly.

Convolutional layers

The input for the convolution2D layer is the NumPy array, some of the parameters include the number of filters, size of filters, size of input images, etc. The filter collection is weights that are randomly produced. To construct the feature image, the filters of size (3,3) are applied to the image with the weights. In total, we have used 4 convolutional layers for expression classification and 3 convolutional layers for age and gender classification. Convolution produces feature

maps that show how pixel values are improved.

Pooling is a dimension reduction technique that is commonly used after one or more convolutional layers have been applied. When developing CNNs, this is a crucial stage because adding more convolutional layers can significantly increase processing time. We utilized the MaxPooling2D pooling approach, which uses (2, 2) windows across the feature map to maintain only the maximum pixel value and down-size image by 4.

Dense layers

Dense layers accept a large number of input features and transform them using trainable weights and layers. Forward propagation of training data is used to train these weights, followed by backward propagation of errors. By adjusting hyper-parameters like learning rate, we may regulate the training pace and complexity of the architecture.

Dropout is one way to avoid over-fitting and generalization on previously unknown data. During training, Dropout randomly picks a subset of nodes (typically less than 50 percent) and sets their weights to zero. This strategy efficiently controls the model's noise sensitivity during training while retaining the architecture's required complexity.

Output layer

At the output layer, we employed softmax instead of the sigmoid activation function. For each emotion class, this output appears as a likelihood. As a result, the model can display the probability composition of the emotions in the face in great detail. Finally, each image is classified into one of the seven basic expressions based on the emotion which has the highest probability, one of the two genders, and one of eight classes of ages based on which has the highest probability.

The output layer, which gives the probability of each label, picks the highest probability label.

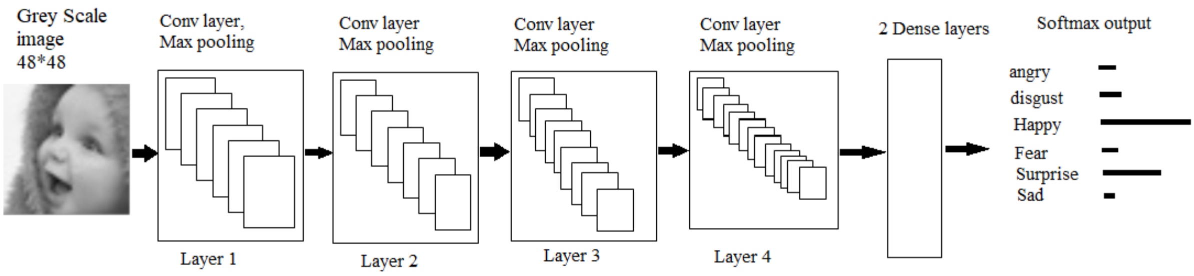

Figure 5.14: Working model of expression

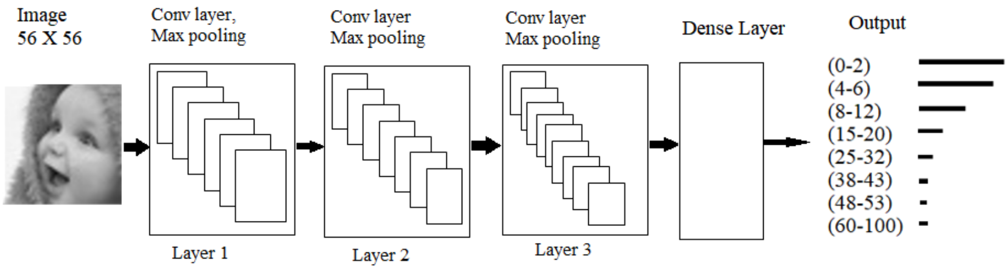

Figure 5.15: Working model of age

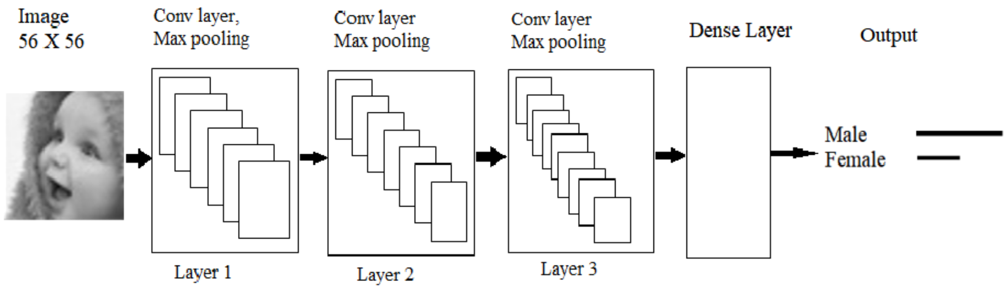

Figure 5.16: Working model of gender

5.5 Result Analysis

Performance The final CNN turned out to have a validation accuracy of 71%. This makes a great deal of sense. Because our expressions frequently contain a mix of emotions, it might be difficult to represent an expression with only one label. When a model predicts an emotion inaccurately, the correct label is frequently the second most likely feeling.

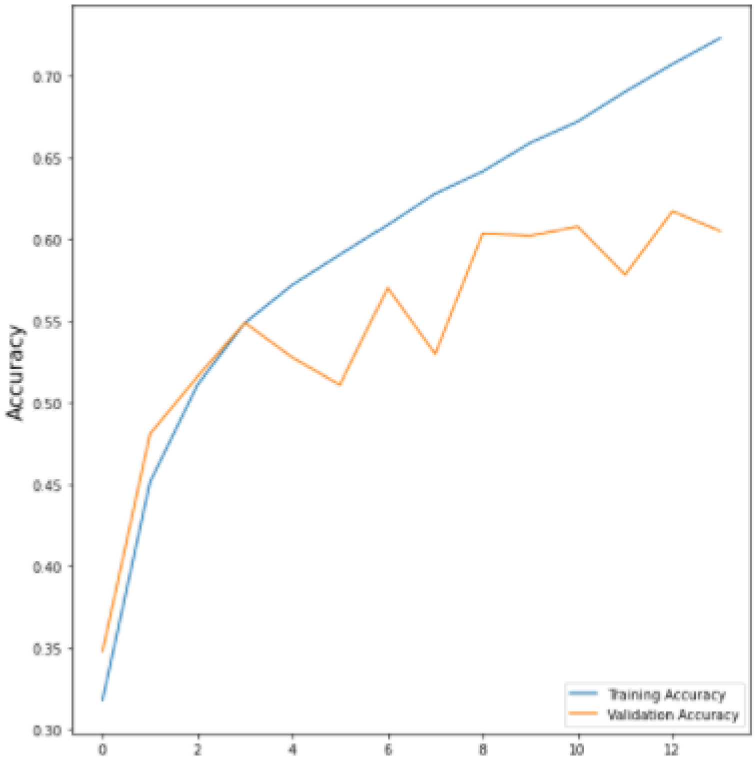

Figure 5.17: Accuracy

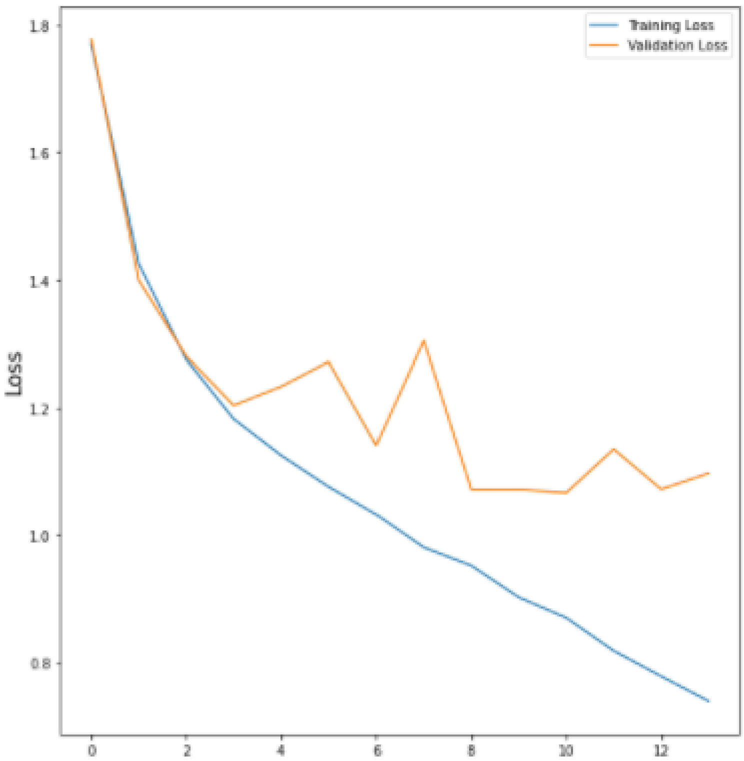

Figure 5.18: Loss

CHAPTER 6

Testing and Results

6.1 Introduction

Our model downsizes the training and testing images into 48x48 for expression and 56X56 for age and gender and then process. The model uses 4 convolutional layers and uses a haar cascade algorithm to detect the face and output the probabilities of all the seven expressions,two for gender and 8 for age and pick up the one with the highest probability for expression, age and gender .
Below are the steps used to lively detect the image through a web camera [17],[18].

1. Download the haarcascade_frontalface_default.xml and use the location where that is downloaded and store it in a variable (face_classifier).

2. Use the location where the .h5 file created after running the model and store in a variable(classifier).

3. Use the openCV method VideoCapture(0) by cv2.VideoCapture, 0 is to capture from web camera of your currently working PC.

4. Draw a rectangle over the face

5. Get the label based on the emotion which gives the highest prop ability.

6. Put the text on the rectangle drawn

7. Press 'q' to exit.

6.2 Design of Test cases and Scenarios

Unit Test

Unit test is to test the modules individually. The individual modules of my project are testing the website, testing the age model, testing gender model, testing of expression model.

As shown in below images after testing individual modules, all the modules are working correctly as expected.

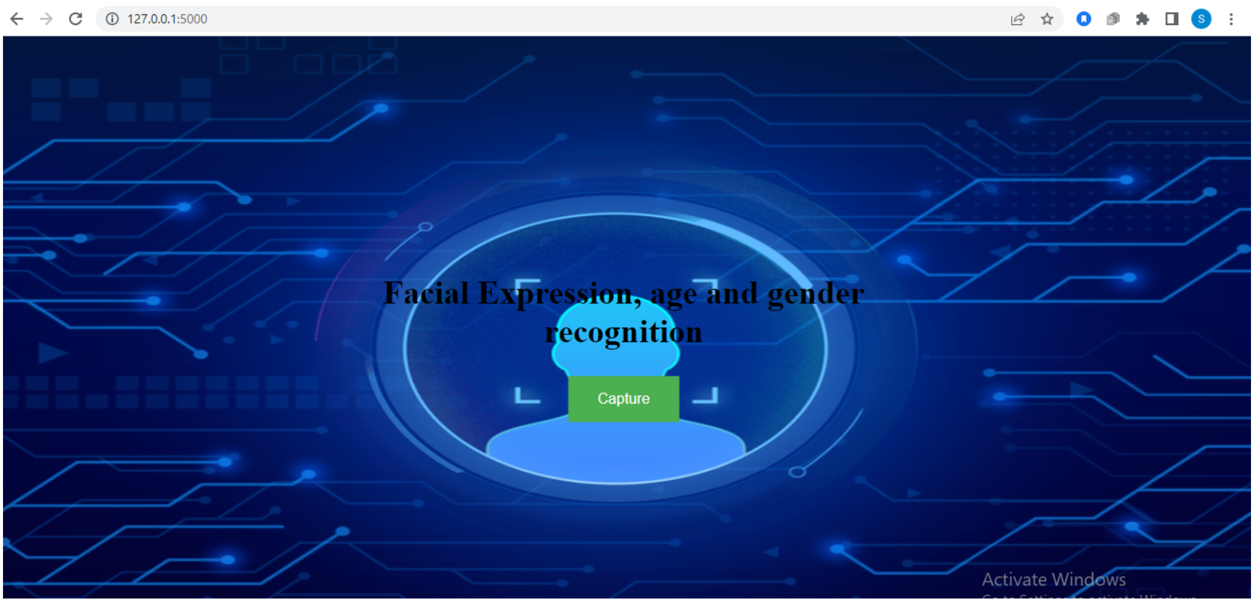

Figure 6.1: Front page of website

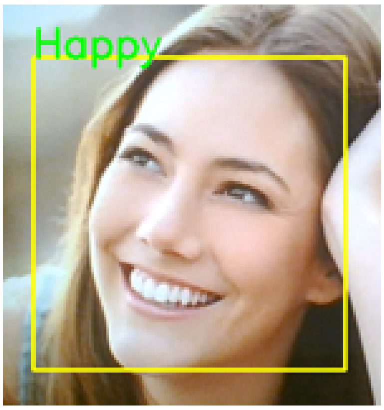

Figure 6.2: Test case for expression

Integration test

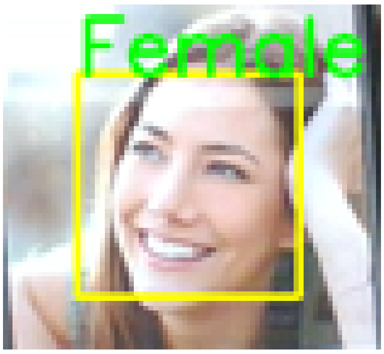

Figure 6.3: Test case for gender

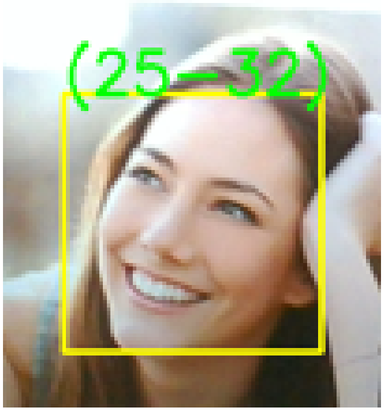

Figure 6.4: Test case for age

Integration test is to check if the model is working correctly as predicted after integrating all the modules.

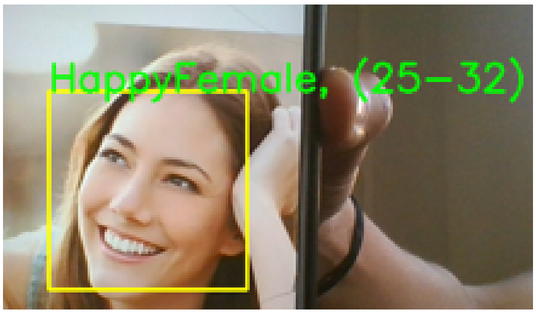

Figure 6.5: Data flow diagram for building model

CHAPTER 7

Conclusion and future enhancements

When a model incorrectly predicts an emotion, gender, or age, the right label is often the second closest emotion. The qualities of the human face are related to geometrical structures that are rebuilt as the recognition system's basic matching template, which are significant to varied expressions.

Future enhancements

It's critical to realize that there isn't a one-size-fits-all approach to building a neural network that will always work. Various issues will need a diverse network architecture and a great deal of trial and error in order to reach acceptable validation accuracy. This is why "black box algorithms" are often used to describe neural networks. In this experiment, we achieved an accuracy of around 70%, which is not terrible when compared to earlier models. However, there are several areas where we need to improve, such as

1. The arrangement of thick layers
2. The percentage of dropouts in dense layers.
3. The count of convolutional layers can be increased.

To improve the model's accuracy, we'd like to add new databases to the system. This project may be expanded to accept the input file as an image and classify it into one of the seven expressions, one of the two genders, and one of the eight age groups.

REFERENCES

[1] John Chris Kwong, Felan Garcia, Patricia Angela Abu, and Rosula Reyes. "Emotion Recognition via Facial Expression: Utilization of Numerous Feature Descriptors in Different Machine Learning Algorithms". In: Oct. 2018. DOI: 10.1109/TENCON.2018.8650192.

[2] Heechul Jung, Sihaeng Lee, Sunjeong Park, Byungju Kim, Junmo Kim, Injae Lee, and Chunghyun Ahn. "Development of deep learning-based facial expression recognition system". In: *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*. 2015, pp. 1–4. DOI: 10.1109/FCV.2015.7103729.

[3] Manas Sambare. "FER-2013 Dataset". In: *IEEE Access* 8 (2020).

[4] Alfredhhw. "Adience Dataset for age and gender classification Dataset". In: *IEEE Access* 8 (2020).

[5] Chao Qi, Min Li, Qiushi Wang, Huiquan Zhang, Jinling Xing, Zhifan Gao, and Huailing Zhang. "Facial Expressions Recognition Based on Cognition and Mapped Binary Patterns". In: *IEEE Access* 6 (2018), pp. 18795–18803. DOI: 10.1109/ACCESS.2018.2816044.

[6] Ghulam Ali, Amjad Ali, Farman Ali, Umar Draz, Fiaz Majeed, Sana Yasin, Tariq Ali, and Noman Haider. "Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis". In: *IEEE Access* 8 (2020), pp. 134950–134963. DOI: 10.1109/ACCESS.2020.3009908.

[7] Huijuan Zhao and Ping Wang. "A Short Review of Age and Gender Recognition Based on Speech". In: *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. 2019, pp. 183–185. DOI: 10.1109/BigDataSecurity-HPSC-IDS.2019.00042.

[8] Ferdous Ahmed, A. S. M. Hossain Bari, and Marina L. Gavrilova. "Emotion Recognition From Body Movement". In: *IEEE Access* 8 (2020), pp. 11761–11781. DOI: 10.1109/ACCESS.2019.2963113.

[9] Priyanka Goel and Suneeta Agarwal. "Hybrid Approach of Haar Cascade Classifiers and Geometrical Properties of Facial Features Applied to Illumination Invariant Gender Classification System". In: *2012 International Conference on Computing Sciences*. 2012, pp. 132–136. DOI: 10.1109/ICCS.2012.40.

[10] *What is a convolutional neural network?* URL: https://www.nvidia.com/en-us/glossary/data-science/convolutional-neural-network/.

[11] *Convolutional Neural Network*. 2019. URL: https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network.

[12] By: IBM Cloud Education. *What are convolutional neural networks?* URL: https://www.ibm.com/cloud/learn/convolutional-neural-networks.

[13] Youhui Tian. "Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm". In: *IEEE Access* 8 (2020), pp. 125731–125744. DOI: 10.1109/ACCESS.2020.3006097.

[14] *OpenCV Haar Cascade: Guide to opencv haar cascade*. 2021. URL: https://www.educba.com/opencv-haar-cascade/.

[15] Mahmoud Harmouch. *Haar cascade classifiers in opencv explained visually.* 2020. URL: https://medium.com/swlh/haar-cascade-classifiers-in-opencv-explained-visually-f608086fc42c.

[16] Aditya Mittal. *Haar Cascades, Explained*. 2021. URL: https://medium.com/analytics-vidhya/haar-cascades-explained-38210e57970d.

[17] Akshitmadan. *Emotion classification: CNN using Keras*. 2021. URL: https://www.kaggle.com/akshitmadan/emotion-classification-cnn-using-keras.

[18] Gil Levi and Tal Hassner. "Age and Gender Classification Using Convolutional Neural Networks". In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*. 2015. URL: \url{https://osnathassner.github.io/talhassner/projects/cnn_agegender}.