University of Mumbai

MSC IT – PART II

Submitted By

MAKRAND PRAFUL GHAG

APPLICATION ID: 93669
SEAT NO: 2500084

SEMESTER: III

Technical Writing & Entrepreneurship Development

Project Documentation

ACADEMIC YEAR : 2023-2024

Centre of Distance and Online Education

DR. SHANKAR DAYAL SHARMA BHAVAN, IDOL BUILDING,

VIDYANAGRI, SANTACRUZ (E), MUMBAI –400098

FACIAL EMOTIONS RECOGNITION

A Project Report

Submitted in partial fulfillment of the Requirements for the award of the Degree of

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)

By

MAKRAND PRAFUL GHAG

Application Id – 93669

Seat No - 2500084

Under the esteemed guidance

Professor Hina Mahmood

DEPARTMENT OF INFORMATION TECHNOLOGY

(Affiliated to University of Mumbai)

Centre of Distance and Online Education

DR. SHANKAR DAYAL SHARMA BHAVAN, IDOL BUILDING,VIDYANAGRI,

SANTACRUZ (E), MUMBAI – 400098

PROFORMA FOR THE APPROVAL PROJECT PROPOSAL

PRN No: 2022027900069422 Application ID: 93669

1. Name of the Student :- Makrand Praful Ghag
2. Title of the Project :- Facial Emotions Recognition
3. Name of the Guide :- Professor Hina Mahmood
4. Teaching experience of the Guide :-
5. Is this your first submission □ Yes, □ No

Signature of the Student Signature of the Course Faculty

Date: ………………... Date: …………………

University Of Mumbai

Centre of Distance and Online Learning

Dr.Shankar Dayal Sharama Bhavan, Kalina, Vidanagari,
Santacruz (E), Mumbai-400 098.

Certificate

This is to certify that the project entitled, "**Facial Emotions Recognition**", is bonified work of Mr. **Makrand Praful Ghag**, Application ID: **93669**, Seat No: **2500084** submitted in partial fulfillment of the requirements for the award of degree of Master of Science in Information Technology from University of Mumbai in the academic year 2023-2024.

Section I ________________

Section II ________________

______________________________ ______________________________

MSc (IT) Co-ordinator, IDOL External Examiner

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this project. Their support, guidance, and encouragement have been invaluable throughout the journey.

First and foremost, I extend my heartfelt appreciation to **Professor Hina Mahmood**, my supervisor, for their unwavering support, insightful feedback, and continuous encouragement. Their expertise and guidance have been instrumental in shaping this project.

I am deeply thankful to **Rizvi College of Science Arts & Commerce, Bandra**, for providing their cooperation and assistance in facilitating access to crucial data and resources essential for this project.

I am indebted to my family for their unwavering support, understanding, and patience throughout this endeavor. Their encouragement and belief in my abilities have been a constant source of motivation.

Furthermore, I would like to express my gratitude to my friends and peers for their encouragement, insightful discussions, and constructive criticism, which have enriched this project.

Finally, I would like to acknowledge the contributions of all the individuals who, directly or indirectly, played a part in this project.

Thank you all for your support and encouragement.

DECLARATION

I am MAKRAND PRAFUL GHAG, a student of MSc.IT wish to state that the work embodied in this project titled **“FACIAL EMOTIONS RECOGNITION”** forms my own contribution to the project work carried out under the guidance of Prof**.** at the department of Information Technology. This work has not been submitted for any other degree of this or any other university. Wherever references have been made to previous work of other, it has been clearly indicated as such andincluded in the bibliography.

Signature of candidate

MAKRAND PRAFUL GHAG

Certified by -

Signature of guide

Date: -

ABSTRACT

Facial emotions recognition is one of the main applications of computer vision and Machine Learning. It contains a lot of information Visually rather than articulately. Automatic facial expression recognition system has lot of applications like human behavior understanding, detection of mental disorders, and synthetic human expressions.

To recognize the facial expression with higher recognition rate is still a major task. In the past, many models have been developed to detect the expressions with higher accuracy. But none of the models have the peak point of accuracy. But with the advancement of internet lot of data is getting generated every day. It has given hope to increase the accuracy of the model through data centric approach. I have covered this project chapter wise.

In chapter 1, I have discussed the introduction of facial expression recognition which basically contains seven basic human expressions. I have also covered the motivation behind this project. I have formulated the problem statement.

In chapter 2, I have discussed the literature review followed by basic four phases of facial expression recognition process. I have covered various approaches to make the model of facial expression recognition. In this chapter, I have mentioned the comparison of accuracy of different datasets.

In chapter 3, I have discussed the analysis & requirements of my project. I have covered the details of the dataset, library, and packages used, algorithm of the project, software & hardware requirement etc.

In chapter 4, I have represented the details about face registration process that includes haar features, adaboost, cascading, etc.

CONTENTS

LIST OF IMAGES

LIST OF TABLES

CHAPTER 1

1.1 INTRODUCTION

In the present age of computers, smart phones, high speed internet connectivity, the modern technology has extended its limits and there are no boundaries for the scope of development which results in fulfilling of actual desires. In the last decade, a lot of research work has been done in the field of digital image and image processing and it is still going on to reach to its maximum peak of efficiency.

Image processing: The field of processing the signals where the input and output signals are images. At present, its application is getting used wide spready.

Facial emotion recognition is one of the most important applications of image processing. The human emotions can be identified by the expressions on their face. The main importance of facial expression is it acts as a key role in interpersonal communication. In Artificial Intelligence and robotics, automatic recognition of facial expression is one of its applications which is important for the upcoming future technologies.

Various Applications of facial expression recognition are as follows:

- Personal identification and access control.
- Videophone and teleconferencing.
- Forensic application.
- Human-computer interaction.
- Automated Surveillance.
- Cosmetology and so on.

Human facial expressions can be categorized into following categories:

1. Neutral
2. Angry
3. Disgust
4. Fear
5. Happy
6. Sadness
7. Surprise

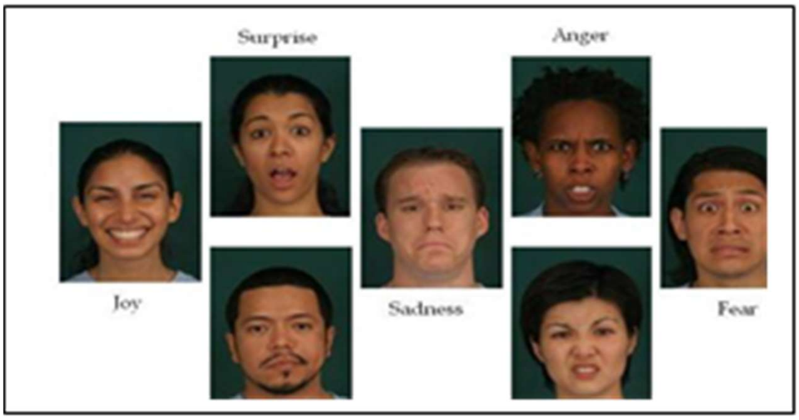


Fig.1.1

The main objective in this project is to train the model by providing it the large number of pictures containing human expressions and make the model to learn those expressions. Then make the model to identify the expression which is based on previous learning. Many projects have already been done in this area. Main aim is to make this project with improved accuracy.

1.2 MOTIVATION

The motivation behind this project are as follows:

- This project can help in providing hints about stress or depression of particular person through recognition of his expression.

Fig.1.2

- Various debates took place regarding the emotions display by the world-famous masterpiece of Mona Lisa. Experts claims that the masterpiece is a mixture of various emotions such as 83% happy, 9% disgust, 6% fearful, 2% angry.

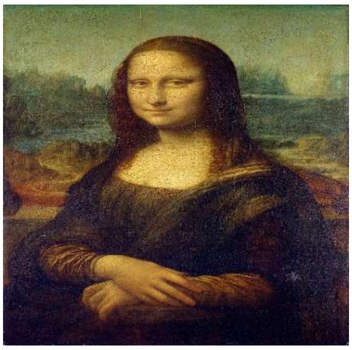


Fig.1.3

- Another aspect for motivation we get in that the big companies and MNC and various working organization can be predicting the emotion of their employee on a particular working day. So that the efficiency of the employees can be improved.

1.3 PROBLEM DEFINITION

As stated earlier, human facial expressions are the combination of seven basic emotions that is happy, sad, surprise, fear, anger, disgust and neutral. They are sometimes under stable, sometimes combination of complex signals in an expression that holds big quantity of information about the state of our mind. For instance, Retailers may use these parameters to analyze their customer's interest int the products if they wish. Doctors in healthcare sector can give better services by using these additional parameters during the patient's treatment. Humans are expert in identifying the emotions of other. Even a small kid can do this. But the main question that comes in our mind in that a computer can do better work than us in identifying? To answer this question, I have designed a deep learning neural network that provides machine the ability to predict of our emotional states.

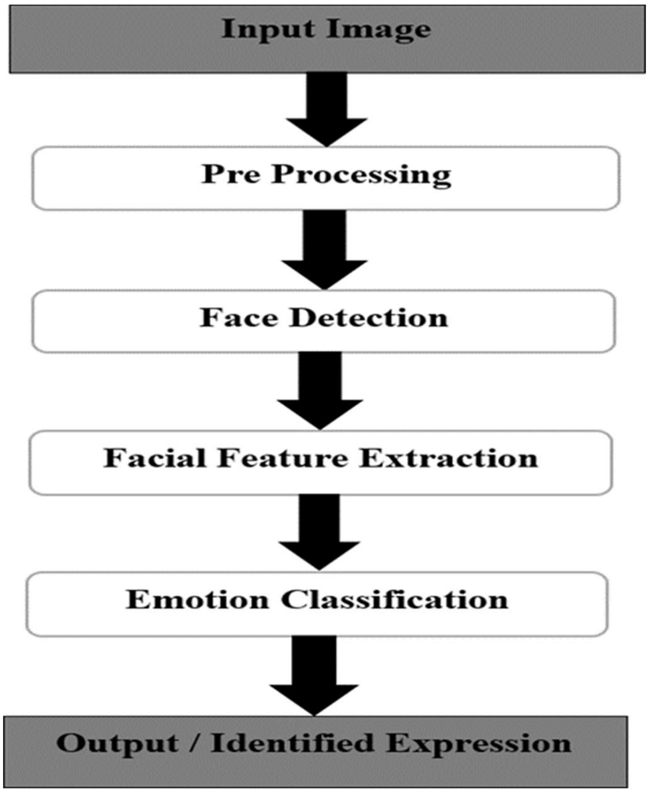


Fig.1.4

CHAPTER 2

LITERATURE REVIEW

On the basis of various literature reviews, I concluded that to implement this project, need to perform four basic steps.

1. Preprocessing
2. Face Recognition
3. Facial Feature Extraction
4. Emotion classification

Preprocessing [8]:

Preprocessing is the terminology that is used to improve quality of images at the lowest level of abstraction where both input and output are intensity images.

The aim of the preprocessing is:

- Reduce the noise.
- Convert the image to binary / grayscale
- Pixel Brightness Transformation
- Geometric Transformation

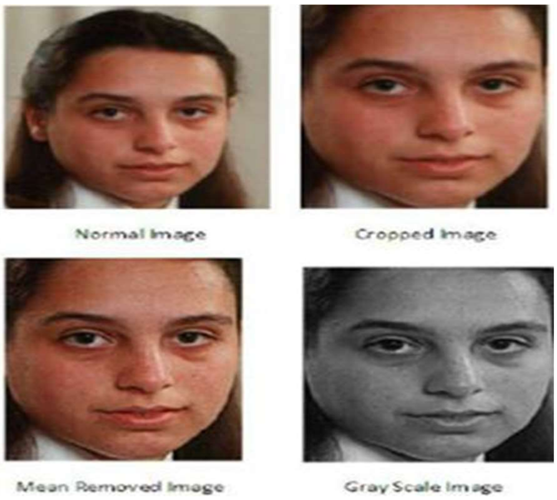


Fig.2.1

Face registration [8]:

It is the process of identifying human faces in the given digital images. In this step, the faces which are in digital image get located by the help of landmark points called “face localization “or “face detection”. Then the detected faces will go through geometrical normalization which later matches to image template called “face registration”.

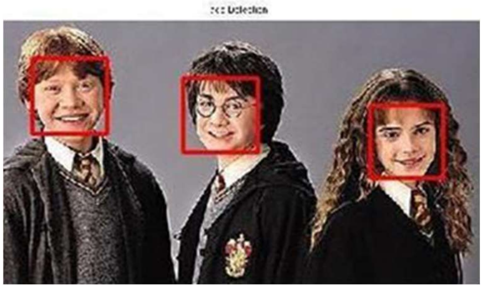


Fig.2.2

Facial Feature Extraction [8]:

It is a crucial step in facial expression recognition. It is defined as the process of track down points, specific regions, landmarks, curves etc. in as 2-D image or a 3-D range image. In this step, a geometrical feature vector is obtained from the registered image. The feature that can be extracted are:

- Lips
- Eyes
- Eyebrows
- Nose tips

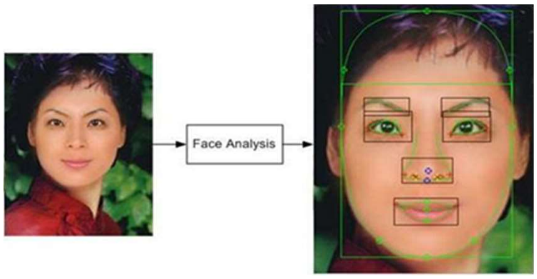


Fig.2.3

Emotion classification [8]:

In this step, the model attempts to identify one of the seven basic emotions from the given faces. Facial expression recognition can be implemented by various approaches. They are as follows:

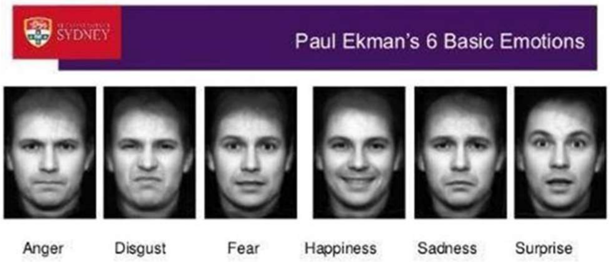


Fig.2.4

Facial expression recognition can be implemented by various approaches and they are as follows:

2.1 NEURAL NETWORK APPROACH [2]

The neural network consists of a hidden layer of neurons. It contains images which maps to the images of similar types also known as local image sampling. It also contains self-organizing map (SOM) neural network and a convolution neural network. The SOM allows mapping of input values from a large set of images to output values of smaller set of images. Convolutional neutral network allows rotation, scale, translation, deformation of images. Its main task is to extract larger features from a hierarchical set of layers.

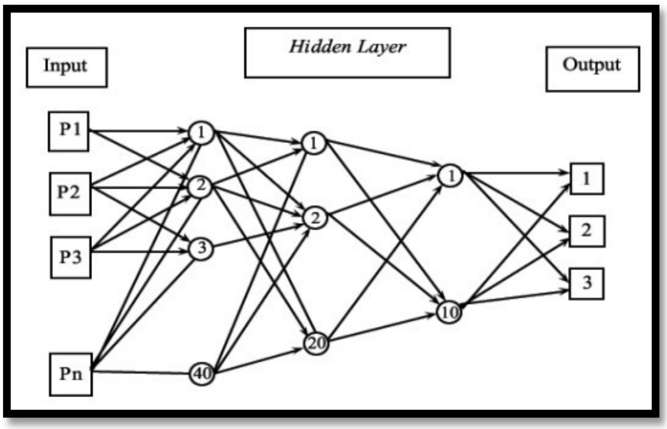


Fig.2.5

2.2 PRINCIPAL COMPONENT ANALYSIS [1]

PCA method works on the principle of eigen space projection. Basically, an image's projection to the matrix is obtained by maximizing the image covariance.

In this, the matrix and the image have correlation in each training data. The 1-D vectors are projected into feature space i.e. the human face would be resembled again after converting it into the matrix of the same size of original face. Principal component analysis can be defined as a method of statistical process that takes an orthogonal transformation to convert a set of observation of all possible correlated variables to a set of values of linearly uncorrelated variable called principal components.

2.3 GABOR FILTER [3]

Gabor filter has been named after Hungarian electrical engineer and physicist Dennis Gabor. It is mainly used for extract features. Basically, it is a kind of linear filter which is largely used for texture analysis. It means for any given particular image it looks out for particular frequency content in the particular path in a localized area around the point or area of analysis. Many vision scientists have claimed that Gabor filters are similar to human visual system. So, we can conclude that the analysis of image is similar as the perception in human visual system.

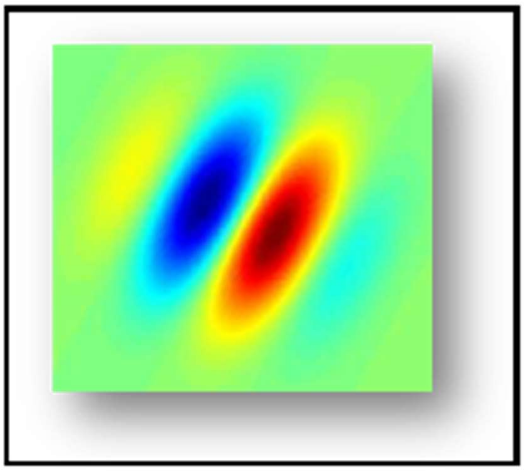


Fig.2.6

2.4 SUPPORT VECTOR MACHINE [4]

Support Vector Machines are generally are supervised learning model that are linked with learning models which are used to analyses data for classification and regression. The main intuition behind support vector machine is to create a hyperplane between the classes of two categories (positive and negative points).

Hyperplane can be defined as the linear line that separates the two classes of negative and positive points. The main point to note that is not only creates a linear line but also two parallel linear planes. When the two classes are separable by hyperplane then it is called linearly separable by hyperplane then it is called nonlinear separable. Support vector machines have supports vectors. Support vectors can be defined as the nearest positive or negative points from which the marginal plane(line) passes through. Support vectors machines have support vector machine kernels that transforms the lower dimension geometry into higher dimension geometry.

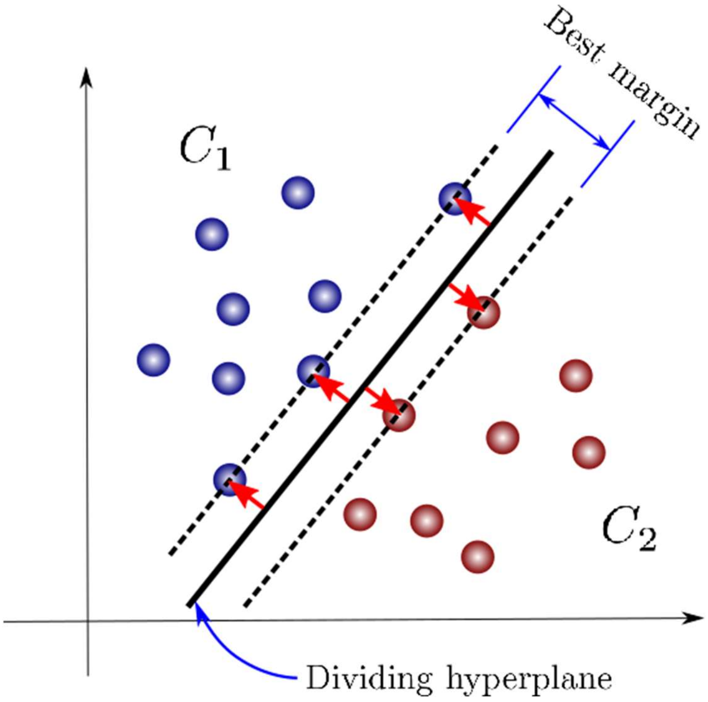


Fig.2.7

2.5 TRAINING AND TESTING DATASET [5]

In the machine learning, training and testing dataset are used by supervised learning models, As in supervised learning models, we want the model to learn from the information we are providing and basis on that information it helps us to predict take decisions by building a mathematical model. So, training and testing data set plays a crucial part in making the model. It is like the food for the model to learn. Generally, multiple datasets are used to build final model. Commonly three datasets are used in different phases of the development of the model.

Training dataset [5]:

Initially, the model is trained by the training dataset that consists of set of examples (weights of connections between neurons in artificial neural networks) of model. It is trained by the process of supervised learning. It contains an input vector and relatively answer vector. Answer vector is denoted as the target. The model runs with the help of training dataset and produces a result. The given result by the model then get compared with the target for every input vector in the training dataset. On the basis of the result being used, the parameters of the models are being adjusted. After the final adjustment of parameters, we can say that the model is fit for use on the validation dataset.

Validation dataset [5]:

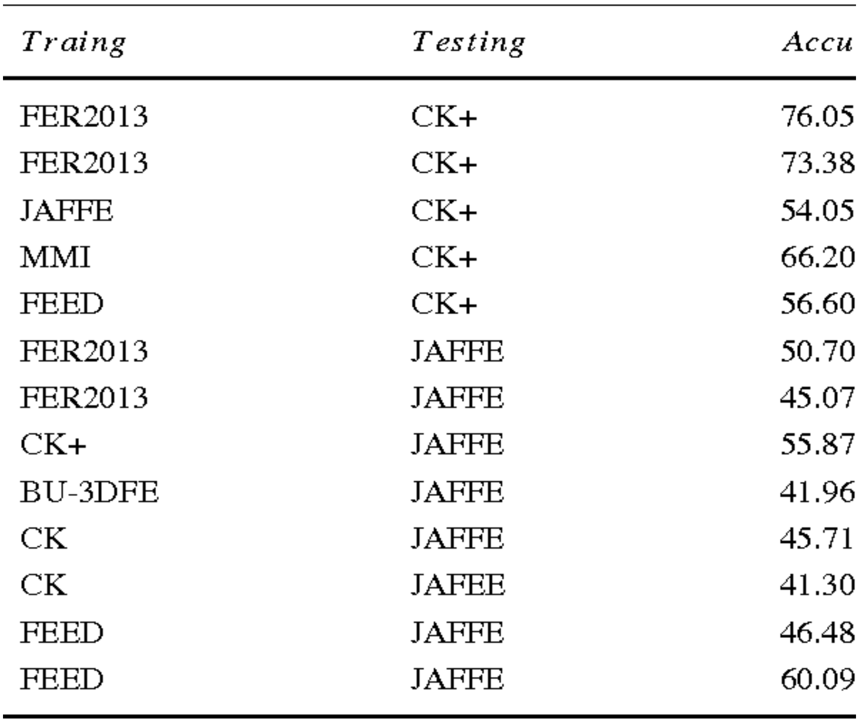
The second dataset which is used to estimate the response for the observation is called validation dataset. It is the way of unbiased evaluation of the fitted model by adjusting the model's hyper parameters. It is used by early stopping. Early stopping is used to stop the training when the errors on the validation datasets gets increases. As it can bet the sign of overfitting of the model which we do not wish to occur. If there are less errors on the validation dataset, then we can consider the training of the model accurately.

Test dataset [5]:

Lastly, the data set which is obtained to produce unbiased evaluation of a final model fit on the training dataset is called the test dataset. Various facial dataset online available are as follows:

- Japanese Female Facial Expression (JAFEE).
- FER
- CMU Multiple
- Lifespan
- MMI
- FEED
- Ck

Accuracy of various datasets:



Traing	*Testing*	*Accu*
FER2013	CK+	76.05
FER2013	CK+	73.38
JAFFE	CK+	54.05
MMI	CK+	66.20
FEED	CK+	56.60
FER2013	JAFFE	50.70
FER2013	JAFFE	45.07
CK+	JAFFE	55.87
BU-3DFE	JAFFE	41.96
CK	JAFFE	45.71
CK	JAFEE	41.30
FEED	JAFFE	46.48
FEED	JAFFE	60.09

Table 1

Accuracy of various approaches are stated as follows:

Sr No.	Method / Technique / Database	Result / Accuracy	Conclusion	Future Work
1	Neural Network + Rough Contour Estimation Routine (RCER) [15] (Own Database)	92.1% Recognition Rate	In this paper, they describe radial basis function network (RBFN) and a multilayer perception (MLP) network.	-
2	Principal Component Analysis [18] (FACE94)	35% Less computation time and 100% Recognition	Useful where larger database and less computational time	They want to repeat their experiment on larger and different databases.
3	PCA + Eigenfaces [19] (CK, JAFFE)	83% Surprise in CK, 83% Happiness in JAFFE, Fear was the most Confused Expression	Compared expression recognition method based on the video sequence, the one based on the static image is more difficult due to the lack of temporal information.	Future work is to develop a facial expression recognition system, which combines body gestures of the user with user facial expressions.
4	2D Gabor filter [22] (Random Images)	12 Gabor Filter bank used to locate Edge	Used for the detection of salient points and the extraction of texture features for image retrieval applications.	They work on adding global and local colour histograms and parameters connected with the shapes of objects within images.
5	Local Gabor Filter + PCA + LDA [23] (JAFFE)	Obtained 97.33% recognition rate with the help of PCA+LDA Features	They conclude that PCA+LDA features partially eliminate sensitivity of illumination.	-
6	PCA + AAM [24] (Image sequences from FG-NET consortium)	The performance ratios are 100 % for Expression recognition from extracted faces.	The computational time and complexity was also very small. Improve the Efficiency	Extend the work to identify the face and it's expressions from 3D images.

Table 2

CHAPTER 3

ANALYSIS & REQUIREMENTS

3.1 DATA SET

The dataset that I have used in this project belongs to an online platform Kaggle. The name of the dataset is Kaggle Facial Expression Recognition Challenge of the year 2013. The dataset contains 48X48 pixel jpg images of faces. All of the jpg images get converted into grayscale to one of the seven categories based on human facial expression. I have given labels to each category. And the labels are as follows:

- 1= Happy
- 2= Sad
- 3= Angry
- 4= Neutral
- 5= Surprise
- 6= Fear
- 7= Disgust

The training dataset contains total 28,709 example of images. The validation dataset contains 3,589 examples of images, the final test set which contributed for the winning of competition contains 3,589 examples of images.

Emotions label in the dataset

- Happy – 8989 images
- Sad – 6077 images
- Angry – 4953 images
- Neutral – 6188 images
- Surprise- 4002 images
- Fear- 5121 images
- Disgust – 547 images

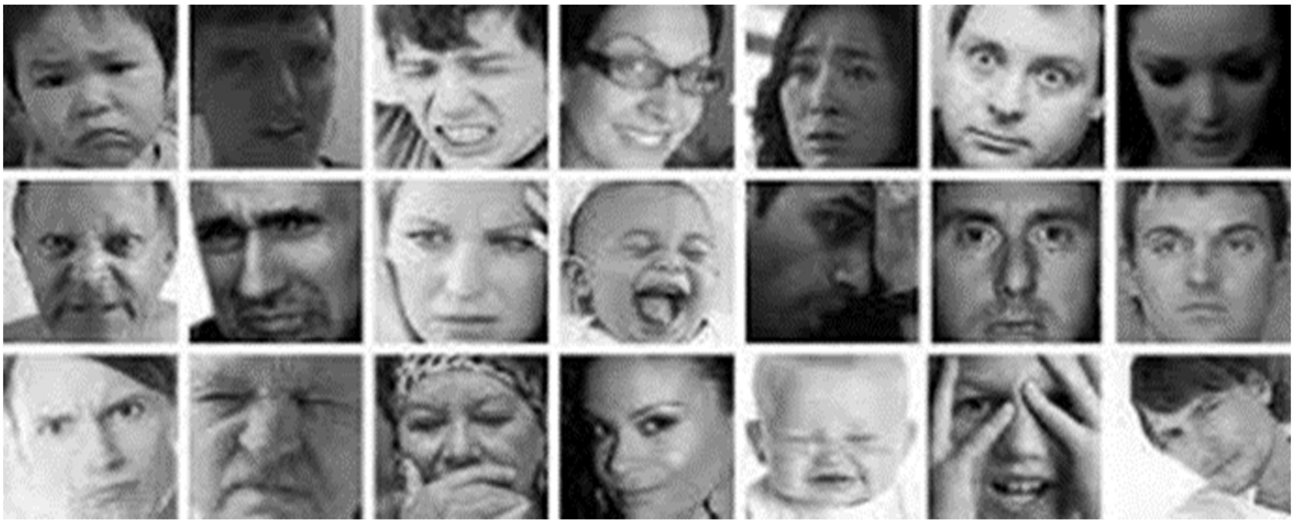


Fig. 3.1

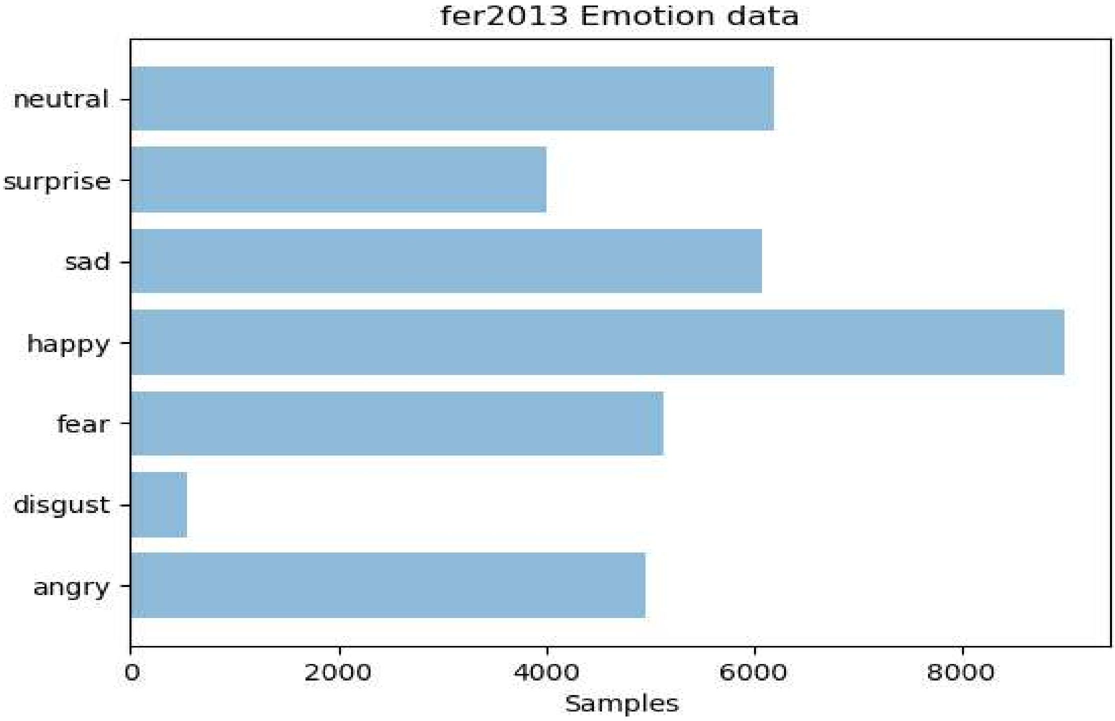

fer2013 Emotion data
neutral
surprise
sad
happy
fear
disgust
angry
0
2000
4000
6000
8000
Samples

Fig.3.2

3.2 Algorithms

Step 1: Collection of a dataset of images.

Step 2: Images will go through preprocessing method.

Step 3: The faces from each digital image will be detected.

Step 4: The jpg images will be converted into grayscale images.

Step 5: The pipeline ensures every image can be fed into the input layer as a (1,48,48) numpy array.

Step 6: The numpy array are getting passed to the convolution 2D layer.

Step 7: Convolution creates feature maps.

Step 8: Pooling method called maxpooling 2D that uses (2,2) windows across the feature map only keeping the maximum pixel value.

Step 9: During the training, neural network forward propagation and Backward propagation are getting performed on the pixel values.

Step 10: The softmax represents the probability for each emotion class.

3.3 Data Flow Diagram

Level 0

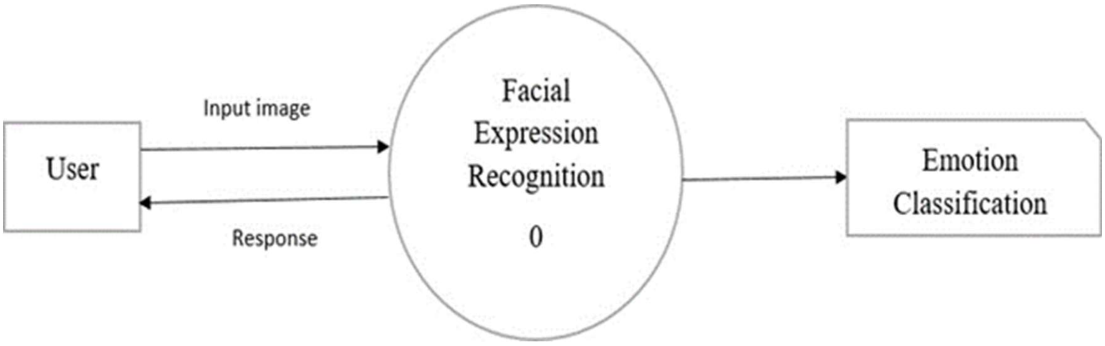


Fig.3.3

Level 1

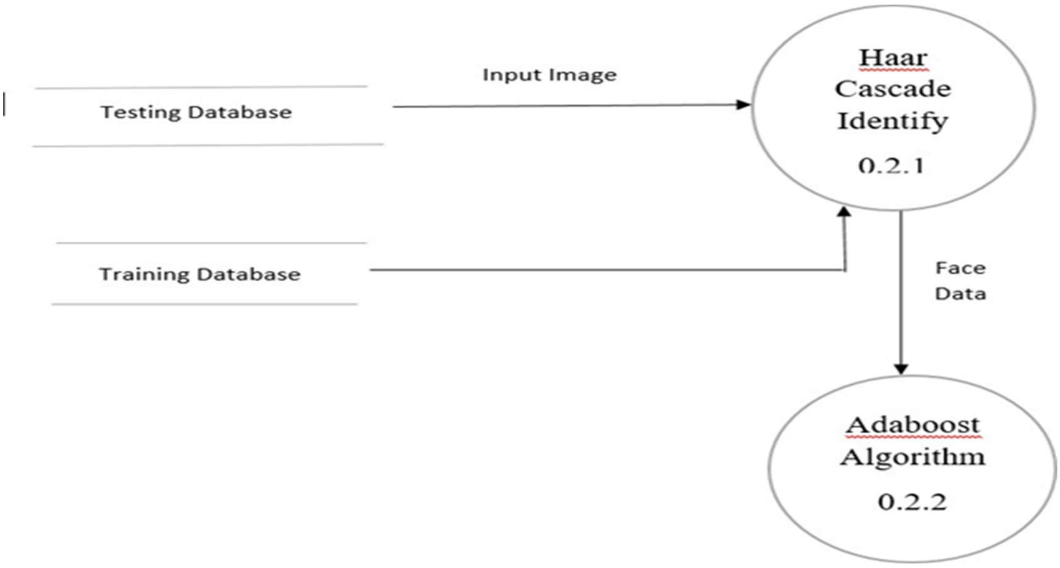


Fig.3.4

Level 2 Face Detection

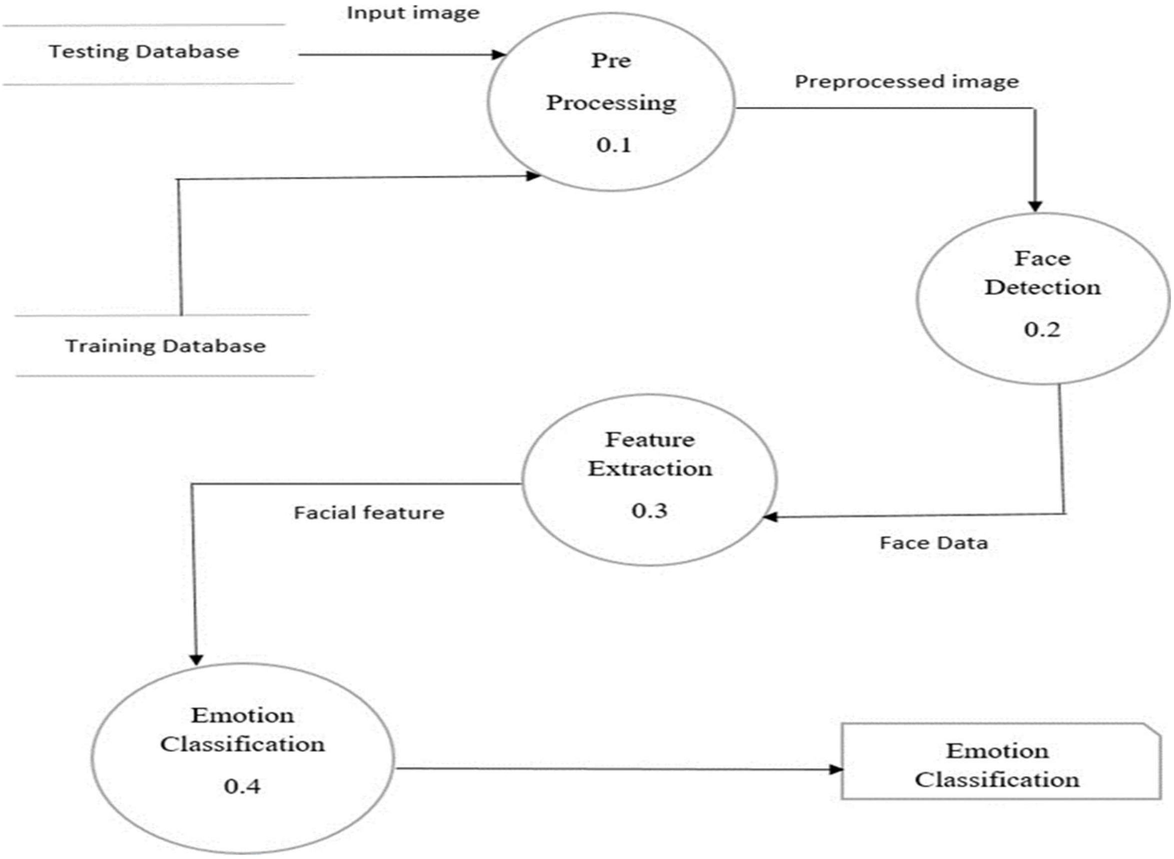


Fig.3.

Emotion Classification

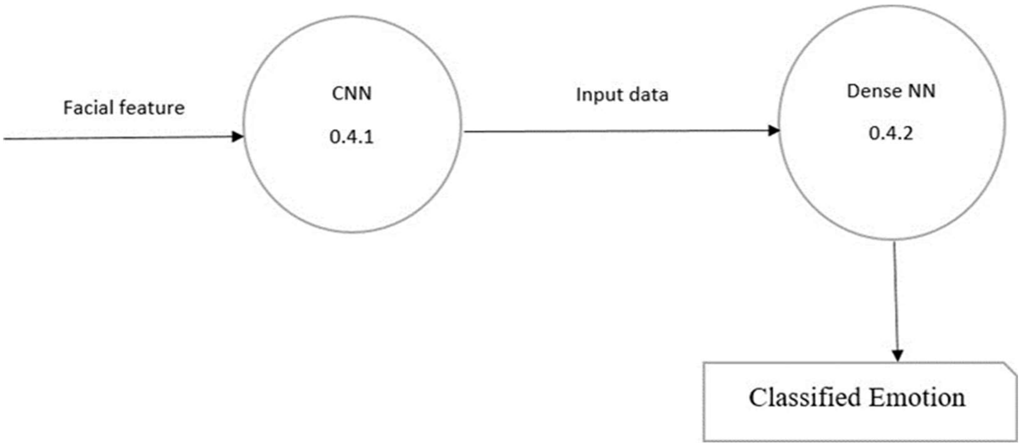


Fig.3.6

3.4 Software & Hardware Requirements

Since, I have developed the project in python, therefore I have used Anaconda for python and spyder.

Anaconda

It is an open source platform for coding in python and R languages. It is beneficial for the work related to data science and machine learning as it allows large scale data processing, predictive analysis, scientific computing.

Spyder

Spyder is an open source platform integrated development environment (IDE) for scientific programming in the python as well as other open source software.

Hardware Requirement

PROCESSOR: Intel core i5 processor with minimum 2.9GHz speed.

RAM: Minimum 4GB.

HARDDISK: Minimum 500 GB.

Chapter 4

PROPOSED METHODOLOGY & APPROACH

4.1 Library and Packages

4.1.1 OpenCV

Open cv (open source computer vision library) is a kind of open source computer vision and machine learning software library. It was developed to provide a common infrastructure for computer vision applications. The library has a collection of greater than 2500 optimized algorithms which consists of both computer vision and machine learning algorithms .These algorithms are helpful in detecting recognized faces, classify human actions in videos, identify objects , track camera movements, track moving objects, extract 30 models of objects, find similar images from an image database and so on . It has C++, python, Java and MATLAB interfaces and supports windows, linux , Android and MacOS.

Application areas of OpenCV are as follows:

- 2D and 3D feature toolkit
- Egomotion estimation
- Facial Recognition system
- Gesture Recognition
- Human- computer Interaction
- Mobile robotics
- Motion understanding
- Object identification
- Segmentation and recognition
- Structure from motion
- Motion tracking
- Augmented reality

4.1.2 Numpy

Numpy is exactly a short term given for “Numeric Python” or “Numerical Python”.

It is basically a module for python that gives quick precompiled functions for mathematical and numerical problems. It is a provider that provides powerful data structures for efficient computation of multidimensional arrays and matrices that makes python programming language even more powerful. Apart from that it also provides large libraries of high-level mathematical functions to operate on these matrices and arrays.

It contains various features such as:

- A powerful N- dimensional array object sophisticated functions
- Tools for integrating C++/C and fortran code.
- Useful linear algebra, Fourier transform etc.

4.1.3 Numpy Array

A numpy array is a kind of array that contains values of all same type which indexed by a tuple of no negative integers. The number of dimensions depends upon the rank of array. The shape of the array depends upon the array along each dimension.

4.1.4 Keras

Keras is a high-level neural network API which is written in python. It has the capacity to run on the top of tensor flow, CNTK or Theano. The aim of the development behind the keras is to be establishing a fast experimentation. It has various implementation that are generally used by neural network building blocks. The implementation of keras that it is providing and layers, objectives, activation functions, optimizer and a host of tools to make the image and text data work easier. It also allows its users to develop their deep models on smart phones, on the web, or on JVM (java virtual machine). It also allows its users to train their dataset using deep learning models on clusters of GRAPHIC PROCESSING UNITS(GPU).

4.1.5 Tensor Flow

It is a library for python programming language for fast mathematical computing which was developed and released by Google. It was developed for the purpose to allow the users to create deep learning models directly.

4.1.6 Sigmoid Function

The main work of sigmoid function is that it produces the end point(activation) of inputs multiplied by their corresponding weights in a neural network. Let us assume we have two columns (features) of input data and one hidden node(neuron) in our neural network. Then each feature would be multiplied by its relative weight value and then get added together. Then the addition of features and weights are passed through sigmoid (logistic regression).

To convert this assumption into neural network we need to add more hidden units.to add hidden units, we need to provide a path from every input feature to those hidden units. Then everytime these features will be multiplied by their corresponding weights and sum of features and corresponding multiplied weights will add up and will get pass through sigmoid which will result in unit's activation.

Properties of Sigmoid Function

- Sigmoid function returns a real valued output
- Sigmoid function takes any range of real numbers and returns the output in the range of 0-1
- Non-negative: if a number is a greater than or equal to zero
- Non- positive: if a number is less than or equal to zero

4.1.7 Softmax Function

Softmax function is responsible for calculation of probabilities of each target class over all possible target classes. It can be defined as the function that calculates the probability distribution of the event over 'n' different events. The great advantage of using a softmax function that it returns high probability of targets classes which we desire to achieve. As the range of probabilities is between 0 to 1, and the sum of all probabilities will be equal to one. The method which it uses is it computes exponential of the given input value and the sum of exponential values of all input values.

The output of this function will be resulting in as the ratio of exponential input value to the sum of exponential input value to the sum of exponential of all inputs.

4.2 Face Registration

4.2.1 Haar Features

The main work of haar features is to detect edge and it is similar to kernels. Generally, all the humans have some common features. The features are such as the eye portion is darker than upper check portion, nose area is brighter than eye portion. By the help of this kind of matching features, their location and size of faces helps us to detect the face.

It takes a 24X24 window for the particular image. Every feature is obtained by subtracting sum of pixels under white rectangles from sum of pixel under black rectangles. For every possible size and location of the kernel results in lot of calculations of features. For 24X24 window there would be more than 160000 haar features which is a pretty big quality. To solve this issue, we use integral images.

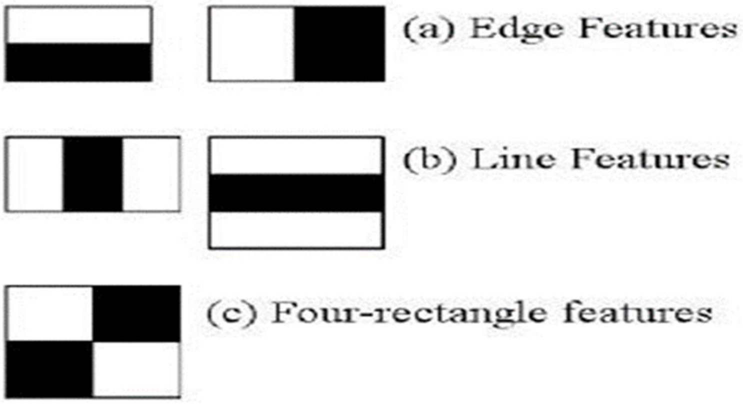
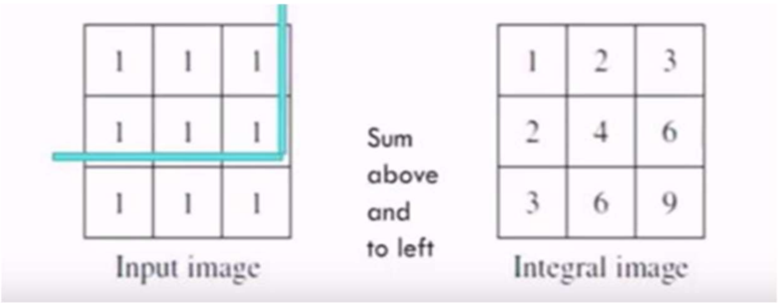


Fig. 4.1

4.2.2 Integral Images

The general idea of integral images is to find the area of pixel. It helps to save time as we do not need to sum up all the pixel values. In fact, we need to use the corner values and then a simple calculation is needed to perform.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'),$$



1	1	1
1	1	1
1	1	1

1	2	3
2	4	6
3	6	9

Fig. 4.2

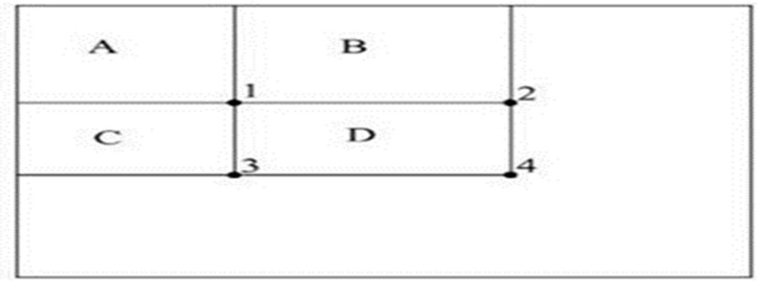


Fig. 4.3

4.2.3 Adaboost

Adaboost is useful for the elimination of unnecessary or unrequired features of haar. A small proportion of these features can be grouped to develop an effectual classifier.
The main task for it is to find these unnecessary features.

For instance, it is not required to find the upper lips as upper lips has more or less constant feature. So, we can eliminate this feature with ease. Adaboost is very useful as it helps us to determine the relevant(necessary) features out of 160000+ features.

Adaboost carries about 2500 classifiers to make a strong classifier to check whether the selected features are relevant or not.

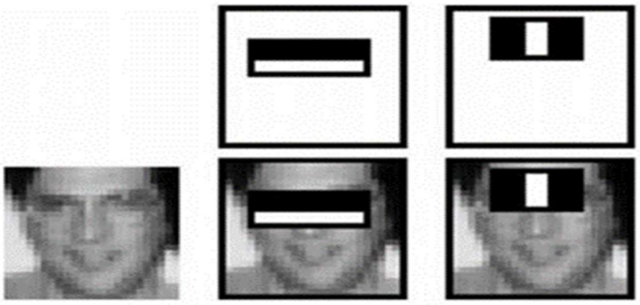


Fig. 4.4

4.2.4 Cascading

Cascading is the method to eliminate non face by evaluating the features from the given set of features. In this, all the features are divided into different groups and are place into different classifiers. Let us assume we have image of 640X480 resolution. We need to pass that image through 24X24 window and from those windows each window will contain 2500 features which is needed to evaluate. Now taking all 2500 features for linear evaluation to decide if it is a face or not. We will better use cascade which will take these features into different classifiers. Out of 2500 features, 1st 10 features are classified in one classifier. Then next 20-30 features into other classifier and next 100 features into another. This method will help us to detect non face early so that it would not require to evaluate again.

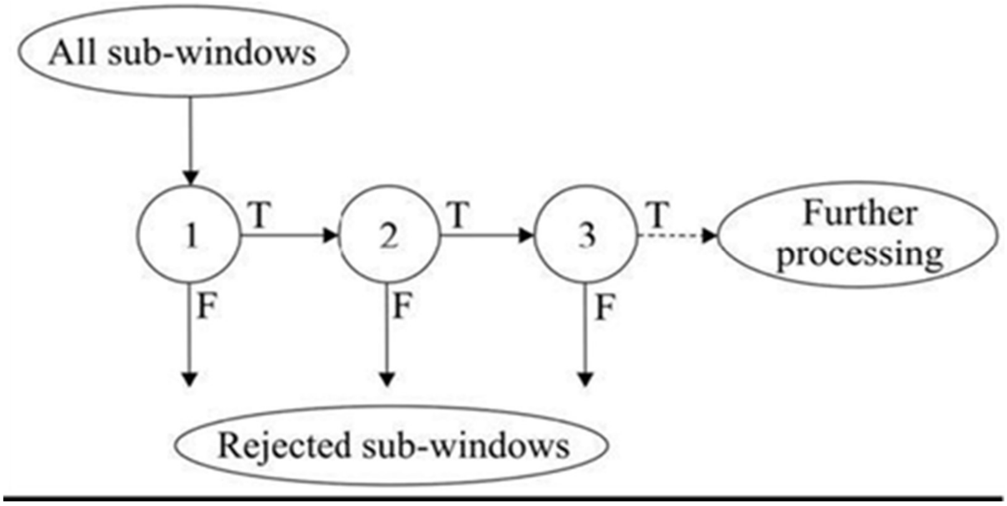


Fig. 4.5

4.2.5 Haar Cascade Classifier in OpenCV

Earlier, we had discussed about the various factors responsible for face registration. We had discussed that we need both positive and negative examples to train the classifier. Then we are required to extract the features. And for that haar features are required to use. We had gone through that to obtain the feature we are required to subtract the sum of pixel under white rectangle from sum of pixel under black rectangle. The summary was it would generate 160000 plus features from a 24X24 windows. To evaluate each feature, we are required to sum of pixels under white and black rectangles. To tackle this problem of 160000 plus features, we introduced integral images.

Integral images are based on the idea of find the area of pixel by using the corner values of pixel. Now we don't need to sum up all pixels under white black rectangles. Then after integral images, we came to know about adaboost whose main task is to eliminate irrelevant features obtained from haar. Adaboost have about 5 combined classifiers to develop a strong classifier.

After adaboost, we came to know about cascading which is responsible to decide the features and provide them to various classifiers. This method of cascading helps us to determine the nonface by providing first 10 features to one classifier, next 20–30 features to other classifier and next 100 features to another classifier and so on.

The combination of all these processes haar features, integral images, adaboost and cascading lead us in the result to final classifier. Final classifier is basically a weak classifier because it contains errors and misclassification. Nut the combination of this classifier with strong classifier leads to provide 95% accuracy in detection of faces. This combination provides a total of 6000 features (reduction from 160000 to 6000 features) which is remarkable. Now if we take an image and take that image through 24X24 window. We will apply only 6000 features to it, we will find whether the image is face or not in little time.

So, this method will help us to detect the face in a single shot and we would not require processing it again. For this, we get a concept of cascade of classifiers which in fact help us to prevent to apply all features to the image. In fact, it tries to detect the face of image in the first stage and so on.

4.3 Deep Convolutional Neural Networks

Convolution Neutral Network can be considered pretty much similar to ordinary neural networks. They contain neurons which have learnable weights and biases.

4.3.1 Overview of DCNN Architecture

DCNN have the architecture in which information flows only in one particular direction, usually from their inputs to their outputs called feedforward networks. CNN (Convolutional neural networks) are inspired biologically to human brains. There are several variations of CNN. But CNN are mainly consisting of convolutional and pooling layers, which are combined into modules. They are present in either one or more fully connected layers, as it should be in a standard feedforward neural network. Modules in CNN are placed in a way such that they are stacked on top of each other to form a deep model. In this deep model, an input is passed through several stages of convolution and pooling. The representation of these operations of convolution and pooling are getting feed to one or more fully connected layer outputs the class label. In the recent years several changes have been made in the architecture of DCNN to the purpose of getting high classification accuracy or reducing computing costs.

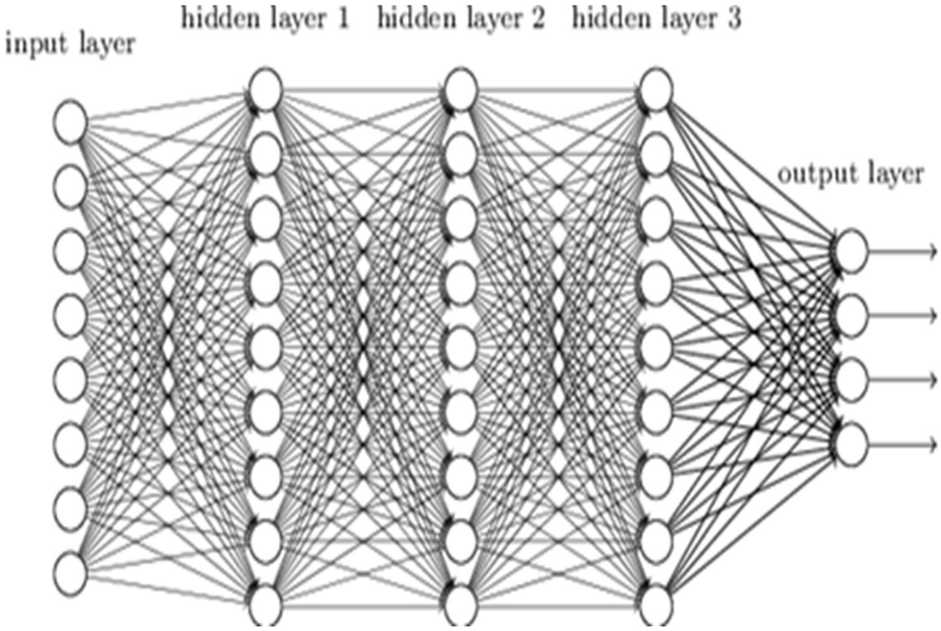


Fig. 4.6

4.3.2 Convolutional Layers

The convolutional layers act as feature extractors, they learn the features representations from their input images. In these convolutional layers, the neurons are placed in feature maps. In feature maps, each neuron has acceptive field (receives input) which is connected to a neighborhood of neurons in the previous layer via set of trainable wights also known as filter bank. The inputs and learned weights are convolved in the way to compute a new feature map. The obtained convolved results are passed through a nonlinear activation function. Different feature maps within the same convolutional layer have different weights so that many features can be extracted at each location.

4.3.3 Pooling Layers

Pooling layers are used as the aim to decrease the spatial resolution of the feature maps and to get spatial variance to input distortions and translations. Earlier average pooling aggregation layers were used to propagate the average of all the input values of a small neighbor hood of an image to the next layer. However, today max pooling aggregation layers are used to propagate the maximum value with in the acceptive fields to the next layer.

4.3.4 Fully Connected Layers

In the fully connected layers, generally various convolutional and pooling layer are piled up in the form of stack on top of each other. This arrangement occurs to pull out symbolic features' representations. These layers are followed by fully connected layers in order to translate these features representation and carry out the functions on the high-level reasoning.

To categorize the problems, we need to use the SoftMax operator on the top of DCNN. Early success can be achieved by radial basis function (RBFs).

If we want to improve the classification accuracy rate, we are required to replace the SoftMax operator by the support vector machine on the top of the classifier.

4.3.5 Training

In training, the CNNs basically uses the algorithms that makes the model to learn. It uses these learning algorithms in order to bring the desired network output by making adjustments in their free parameters. The algorithm that is used commonly behind this aim is "back propagation". It is used to minimize faults by adjusting the whole network's parameters by estimating the gradient of the objective function. One of the most common problems which are generally experienced during the training of CNNs is overfitting. Overfitting can be considered as the poor performance for the test which is held. It is the kind of defect which disturbs the model’s capability to generalize on the unknown data. This problem can be sort out by regularization.

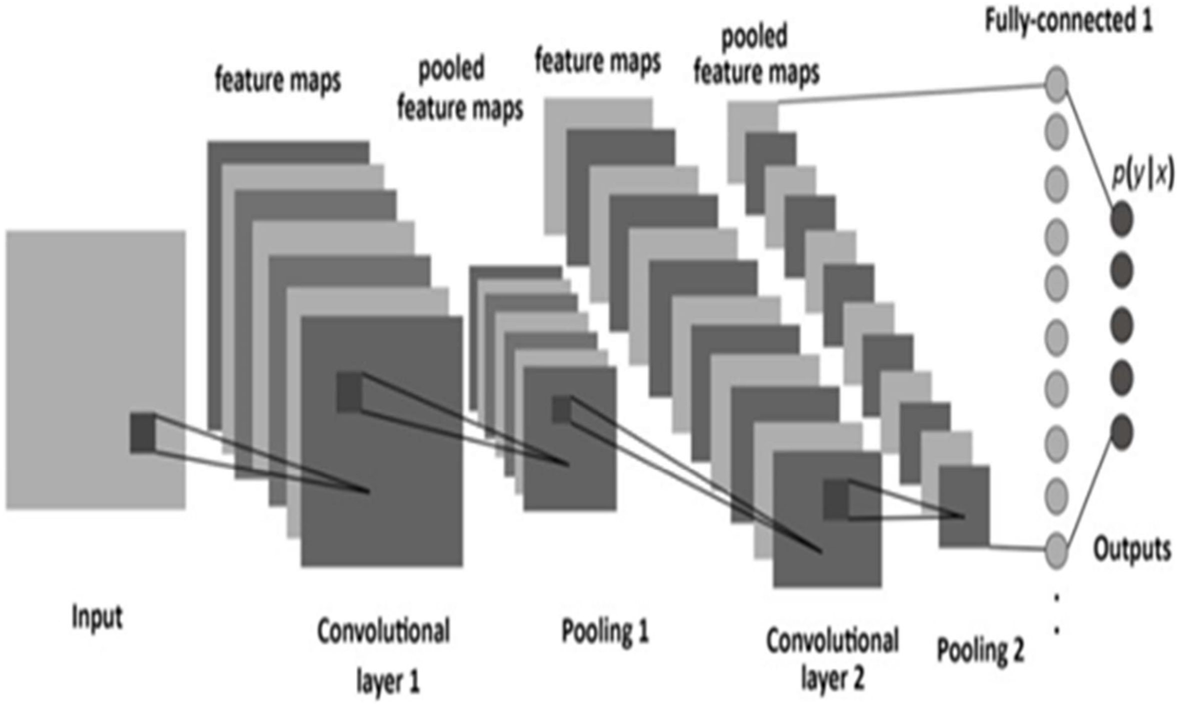


Fig. 4.7

4.4 The Model to Implement the Problem

One of the best ways that can be fruitful for the computer vision is the deep learning technique. For this, convolutional neural networks (CNN) is suggested to take it as a structural blocks to build the model. CNN are based upon the concept that is inspired by the functioning of the human brain to analyze the optical scene. A standard construction of a CNN contains input layer, convolutional layers, dense layers and an output layer. all the layers in CNN construction are arranged in such a manner that they are indexed in sequence like stack layers. The structural model that can be build in the Keras library is to be known as by the name Sequential().

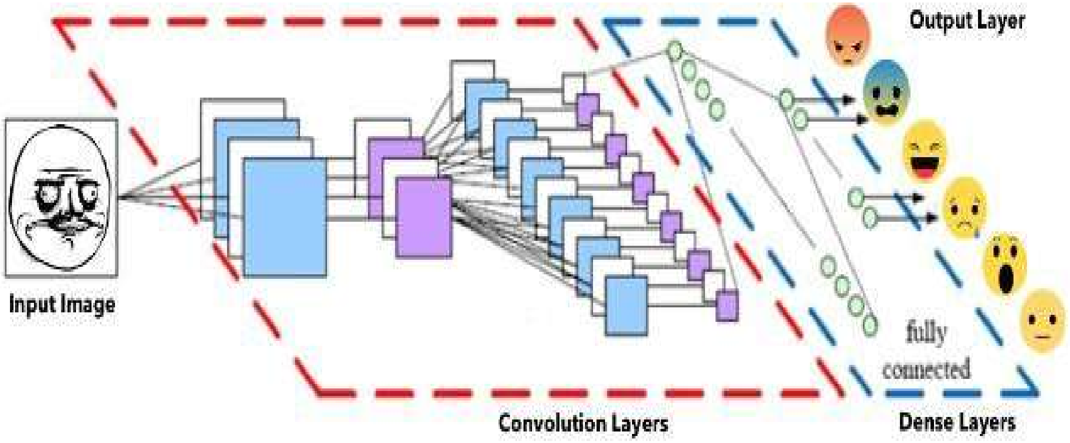


Fig. 4.8

4.4.1 Input Layer

This layer generally carries the images that have pre- arranged set of measures. so, we need to make ensure that the images we are going to pass into this layer should be pre-processed. To ensure that preprocessing has been done that is detection of the facial images, we have used an open source computer vision library called "OpenCV". It contains a default XML file that already has prepared filters and Adaboost in order to make the face cropped and detected.

The cropped face that we have obtained in this layer would be transformed into grayscale with the help of cv2.cvtColor.

It will also get modified by the help of cv2.resize to about 48 by 48 pixels. This step helps in diminishing the measurements if we make comparison of this with the RGB format that contains tri color measurements (3,48,48).

The channel makes sure that every image can be passed to the input image can be passed to the input layer as (1,48,48) NumPy array.

4.4.2 Convolutional Layers

The major function of this layer is to produce the map that contains features of the facial images. The representation of these feature maps shows how the pixel of the image have been improved. To implement this process of producing feature maps. the NumPy array that we have obtained in the input layer have been fed to the convolutional 2D layer. Here, we are required to declare the quantity of filters which can be considered as one of its hyper parameters. The group of filters over here are special for randomly producing weights. Every filter, (3,3) acceptive field covers and slides whole over the area of original image with the shared weights to produce a feature map. A group of feature maps can be obtained by applying various filters one after other in an order.

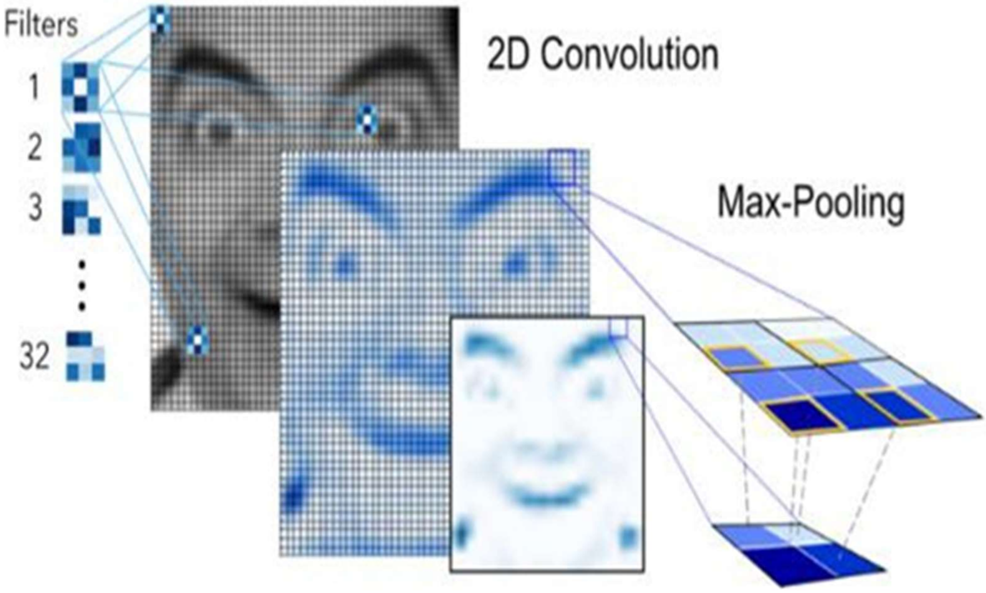


Fig. 4.9

4.4.3 Dense Layers

The concept of the dense layers are influenced by the procedure of the neurons as how they mediate the signal throughout the brain. It is given large number of input attributes and transfigure attributes and transfigure trainable weights. The trainable weights are taught by forward propagation which indulges in the training of data and after that backward propagation which indulges in the propagation of the faults of training data. The function of the back propagation is to estimate the gap between prediction and exact values and need to get back to estimate the weight reconciliation required to every layer before. We are also able to control the speed of training of dataset and the complications of the structure by regulating the hyper- parameters, like the network density and learning rate. As we will increase more to provide the data to the model, our model network layers will be capable to make control slowly unless the faults will be getting minimized.

The greater the number of layers, we will add up to our neural network structure, the better it can grasp the signals.

This procedure will also help us to continuously increase the prevention against overfitting of data training. Another alternative to stop overfitting is the use of dropout. Dropout aimlessly picks up the portion of the nodes and adjust the weights of that portion to zero at the time of training. This procedure helps us to highly control the sensitivity of the model against disturbance (noise) and then ensures the complications of the structure of the model.

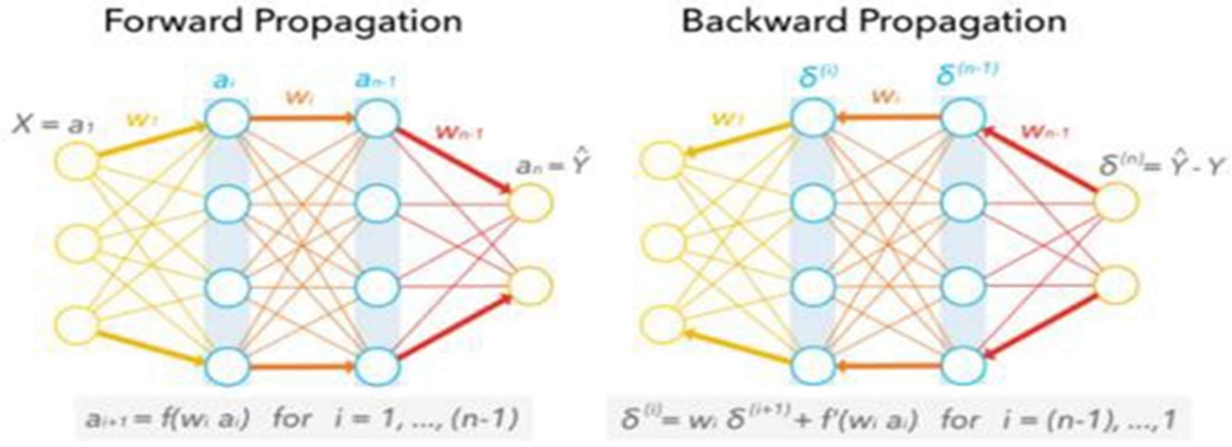

Forward Propagation
Backward Propagation
$X = a_1$
w_1
a_i
w_i
a_{n-1}
w_{n-1}
$a_n = \hat{Y}$
$a_{i+1} = f(w_i\, a_i)$ for $i = 1, \ldots, (n-1)$
$\delta^{(i)}$
$\delta^{(n-1)}$
w_i
w_1
w_{n-1}
$\delta^{(n)} = \hat{Y} - Y$
$\delta^{(i)} = w_i\, \delta^{(i+1)} + f'(w_i\, a_i)$ for $i = (n-1), \ldots, 1$

Fig. 4.10

4.4.4 Output Layer

In this layer, SoftMax function has been used in the project. Basically, this layer is representing solely as a mean of probability for every emotion class. The model can describe the detail probability configuration of the emotions behind the face. I came on conclusion that it is necessary to define (or classify) human facial expressions as complex or mixed expressions because generally our facial expressions are much more complicated as we have thought about on building a simple CNN structure (an input layer, three convolutional layers, one dense and an output later) could lead us to predict the emotions poorly. In fact, I got a low accuracy of about 0.1900 which is guessing of seven emotion with a poor rate. This is why deep learning is extremely important. If we are having complicated facial expressions, then we are required to build a deeper structure so as to get the minute signals.

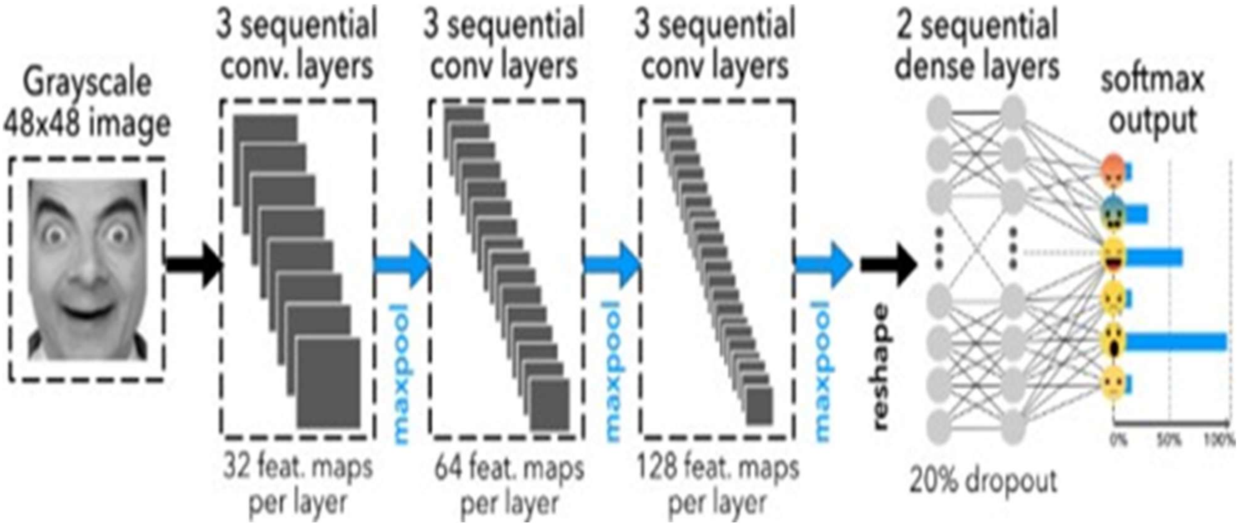


Fig. 4.11

REFERENCES

1. Journal on Facial Expressions recognition Based on Principal Component Analysis (PCA) by Abdelmajid Hassan Mansour, Gafar Zen Alabdeen Salh, Ali Shaif Alhalemi.

2. Journal on Facial Expression Recognition Using Neural Network by Md. Forhad Alia, Mehenag Khatuna and Nakib Aman Turzo.

3. Journal on Facial Expression Recognition Using Log-Gabor Filters and Local Binary Pattern Operators by Seyed Mehdi Lajevardi, Zahir M. Hussain School of Electrical & Computer Engineering, RMIT University, Melbourne, Australia.

4. Facial Expression Recognition Using SVM Classifier by Vasanth P.C, Nataraj K.R.

5. The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi by Lutfiah Zahara (Departmen of Computer Science, Gunadarma University), Purnawarman Musa (Departmen of Computer Science, Gunadarma University).

6.A literature survey on Facial Expression Recognition using Global Features by Vaibhav kumar J. Mistry and Mahesh M. Goyani, International Journal of Engineering and Advanced Technology (IJEAT), April,2013.

7. A Survey on Facial Expression Recognition Techniques by Swati Mishra(Chhattisgarh Swami Vivekanand Technical University, Department of Computer Science & Engineering, Raipur Institute of Technology, Mandir Hasaud, Raipur, Chhattisgarh, India) and Avinash Dhole(Assistant Professor, Department of Computer Science & Engineering, Raipur Institute of Technology, Chhattisgarh Swami Vivekanand and Technical University, Mandir Hasaud, Raipur, Chhattisgarh, India) , International Journal of Science and Research (IJSR),2013 [https://pdfs.semanticscholar.org/e241/25e4e9471a33ba2c7f0979541199caa02f8b.pd f]

8. Recognizing Facial Expressions Using Deep Learning by Alexandru Savoiu Stanford University and James Wong Stanford University.

9. Predicting facial expressions with machine learning algorithms by Alex Young, Andreas Eliasson, Ara Hayrabedian , Lukas Weiss , Utku Ozbulak.

10."Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137–154, 2004.

11. Going Deeper in Facial Expression Recognition using Deep Neural Networks, by Ali Mollahosseini1, David Chan2, and Mohammad H. Mahoor1 Department of Electrical and Computer Engineering, Department of Computer Science, University of Denver, Denver, CO.

12. Journal on Convoluted Neural Network by IIIT, Hyderabad.

13. Journal on Artificial Intellegence by Prof. M.K. Anand, 2014.

14. Journal on Image Processing by Ghuangjhow University.

15. Facial Expression Detection Techniques: Based on Viola and Jones algorithm and Principal Component Analysis by Samiksha Agrawal and Pallavi Khatri, ITM University Gwalior (M.P.), India, 2014.

16. VIOLA JONES FACE DETECTION EXPLAINATION by Rahul Patil.

17. Wikipedia- Artificial Neural Netwok & Convoluted Neural Netwok.

18. Convolutional Neural Networks (CNN) With TensorFlow by Sourav from Edureka
19. Deep Learning Simplified by Sourav from Edureka.