A Project Report

on

Human Emotion Recognition Using Deep Learning From Facial Expression

Submitted in partial fulfillment

requirement for the award of the degree of

Bachelor's of Technology in Computer Science Engineering

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of

Dr. Shobha Tyagi

Associate Professor

Submitted By

Tanmay Gangwar & Pulkit Gangwar

18SCSE1140013 & 18SCSE1140060

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY,GREATER NOIDA

INDIA

DECEMBER,2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the project, entitled **"Human Emotion Recognition Using Deep Learning From Facial Expression"** in partial fulfillment of the requirements for the award of the Bachelors of Technology submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of August, 2021 to December and 2021, under the supervision of **"Dr. Shobha Tyagi"** (Associate Professor), Department of Computer Science and Engineering of School of Computing Science and Engineering , Galgotias University, Greater Noida.

The matter presented in the project has not been submitted by us for the award of any other degree of this or any other places.

Pulkit Gangwar(18SCSE1140060)

Tanmay Gangwar(18SCSE1140013)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. Shobha Tyagi

(Associate Professor)

CERTIFICATE

The Final Project Viva-Voce examination of "**PULKIT GANGWAR(18SCSE1140060) & TANMAY GANGWAR(18SCSE1140013)"** has been held on _____ and his/her work is recommended for the award of B.Tech(CSE).

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: December, 2021

Place: Greater Noida

ACKNOWLEDGEMENT

Apart from the efforts of our, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We would like to show our greatest appreciation to **Dr. Shobha Tyagi** and also our reviewer **Mr. N. Suresh Kumar**. We can't say thank you enough for his/her tremendous support and help. We feel motivated and encouraged every time, We attend his meeting. Without his encouragement and guidance this project would not have materialized.

The guidance and support received from all the members who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Abstract

Human facial emotion recognition plays an important and vital role in the interpersonal relationship. Deep learning (DL) based emotion detection gives much better and exact results than traditional methods with image processing. It has been recognized for a many years and it is a vital topic in the fields of computer vision and machine learning. Artificial Intelligence (AI) system capable of emotion detection through facial expressions. By using Neural Network classifier training, six kinds of different emotional categories are obtained from images. Through this we proposed a Computer Vision (CV) based deep learning architecture for emotion detection from images. We collect the datasets from social media platforms and various websites like kaggle, etc. In this dataset, data will be presented in the form of image type with some facial expressions. The performance of the proposed method is evaluated using datasets for restaurants which are collected from different social media platforms and websites. Our aim to understand the change in emotions of people coming to hotels or restaurants, to make a better relationship with the users.

Keywords:- Artificial Intelligence ,Computer Vision, Neural networks , Deep Learning , Facial recognition and Machine Learning, Supervised Learning.

List of Figure

Table of Contents

CHAPTER-1

Introduction

1.1 Introduction:

The human face is an important part of an individual's body and it especially plays an important role in extraction of an individual's behavior and emotional state. As humans, we classify emotions all the time with/ without knowing it. In spite of their busyness if they see their facial expression then they may be try to do something different. For example, suppose if anyone see that his or her facial expression is happy then he or she will try to be more happier. On the other hand, if anyone see that his or her facial expression is sad then he or she will improve his or her mental condition. Facial expression plays an important role for detecting human emotion. It is a valuable indicator of a person. In a word an expression sends a message to a person about his or her internal feeling. Facial expression is the most important application of image processing. In the present age , a huge research work on the field of image processing. Facial image based emotion detection techniques provides a fast and useful result for emotion detection. The process of recognizing the expression of feelings through facial emotion was an interesting object. After 1960 this topic became more popular, when a list of universal emotion was

established and different system were proposed. Because of the arrival of modern technology our expectation goes high and it has no limitation.

As a result people try to improve this image based emotion detection in different ways. There are six basic universal emotions for human beings. These are happy, sad, angry, fear, disgust and surprise. From human's facial expression we can easily detect this emotion. In this project we will proposed a useful way to detect happy, sad and angry these three emotions from frontal facial emotion. Our aim, which we believe we will develop a method of face emotion detection that is fast, robust, reasonably simple and accurate with a relatively simple and easy to understand algorithms and techniques. The examples provided in this thesis are real-time and taken from our own surroundings.

Image processing is a technique than can convert an image into digital form and perform different kinds of operation on it for getting better image and useful information. Image processing technique used two types of method. These are analog and digital image processing. Analog technique can be used for hard copies and digital technique used for manipulating digital image. The purpose of image processing is divided into five groups. These are: visualization, image sharpening, image retrieval, measurement of pattern, image recognition.

Visualization observe invisible object. Image sharpening that makes better image. Image retrieval finds interesting image. Measurement of pattern measures different objects of an image. Image recognition finds the difference of an image.

1.2 Problem Formulation:

Human facial expressions can be easily classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial emotions are expressed through activation of specific sets of facial muscles. These sometimes subtle, yet complex, signals in an expression often contain an abundant amount of information about our state of mind.

Through facial emotion recognition, we are able to measure the effects that content and services have on the audience/users through an easy and low-cost procedure.

For example:

Retailers may use these metrics to evaluate customer interest.

Healthcare providers can provide better service by using additional information about patient's emotional state during treatment. Entertainment producer scan monitor audience engagement in events to consistently create desired content.

Humans are well-trained in reading the emotions of others, in fact, at just 14 months old, babies can already tell the difference between happy and sad.

But can computers do a better job than us in accessing emotional states?

To answer this question, We are going to design deep learning neural network using computer vision that gives machines the ability to make inferences about our emotional states. In other words, we give them eyes to see what we can see.

Several Projects have already been done in this fields and our goal will not only be to develop a Automatic Facial Expression Recognition System but also improving the accuracy of this system compared to the other available systems.

Fig 1-Types of Emotions

1.2.1 Tools and Technology Used:

❖ SOFTWARE REQUIREMENT :

As the project is developed in python, we have used Anaconda for Python 3.6.5 and Spyder.

1. Anaconda

It is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, Linux, and MacOS.

2. Spyder:-

Spyder (formerly Pydee) is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates NumPy, SciPy, Matplotlib and IPython, as well as other open source software. It is released under the MIT license.

Spyder is extensible with plugins, includes support for interactive tools for data inspection and embeds Python-specific code quality assurance and introspection instruments, such as Pyflakes, Pylint and Rope. It is available cross-platform through Anaconda, on Windows with WinPython and Python (x,y), on macOS through MacPorts, and on major Linux distributions such as Arch Linux, Debian, Fedora, Gentoo Linux, openSUSE and Ubuntu.

❖ Features include:

- editor with syntax highlighting and introspection for code completion
- support for multiple Python consoles (including IPython)
- the ability to explore and edit variables from a GUI

❖ Available plugins include:

- Static Code Analysis with Pylint
- Code Profiling
- Conda Package Manager with Conda

❖ Hardware Interfaces

1. Processor : Intel CORE i5 processor with minimum 2.9 GHz speed.
2. RAM : Minimum 4 GB.

3. Hard Disk : Minimum 500 GB

- Software Interfaces

1. Microsoft Word 2003
2. Database Storage : Microsoft Excel
3. Operating System : Windows10

1.3 Motivation:-

In previous time, for psychologist, analyzing facial expression was an essential part. Nowadays image processing have motivated significantly on research work of automatic face mood detection. There are lots of depressed people lived in our society. Also lots of busy people those who do not know their present mental condition. So we try to develop such an application and by this application they will able to see their present mental condition.

1.4 OBJECTIVES OF THE PROPOSED PROJECT WORK:-

This project is a system which automatically recognizes the emotion represented on a face. Thus a neural network based solution combined with image processing is used in classifying the universal emotions: Happiness, Sadness, Anger, Disgust, Surprise and Fear. Coloured frontal

face images are given as input to the prototype system. After the face is detected, image processing based feature point extraction method is used to extract a set of selected feature points. Finally, a set of values obtained after processing those extracted feature points are given as input to the neural network to recognize the emotion contained. In addition to emotion detection, this project can help detect a few set of objects in the frame too.

1.5 PROJECT FEATURES:-

For automatic FER systems, various types of conventional approaches have been studied. The commonality of these approaches is detecting the face region and extracting geometric features, appearance features, or a hybrid of geometric and appearance features on the target face.

For the geometric features, the relationship between facial components is used to construct a feature vector for training. Two types of geometric features are used based on the position and angle of 52 facial landmark points. First, the angle and Euclidean distance between each pair of landmarks within a frame are calculated, and second, the distance and angles are subtracted from the corresponding distance and angles in the first frame of the video sequence. For the classifier, two

methods are presented, either using multi-class AdaBoost with dynamic time warping, or using a SVM on the boosted feature vectors.

The appearance features are usually extracted from the global face region or different face regions containing different types of information. Unlike a global-feature-based approach, different face regions have different levels of importance. We have to extract region-specific appearance features by dividing the entire face region into domain-specific local regions. Important local regions are determined using an incremental search approach, which results in a reduction of the feature dimensions and an improvement in the recognition accuracy.For hybrid features, some approaches have combined geometric and appearance features to complement the weaknesses of the two approaches and provide even better results in certain cases.

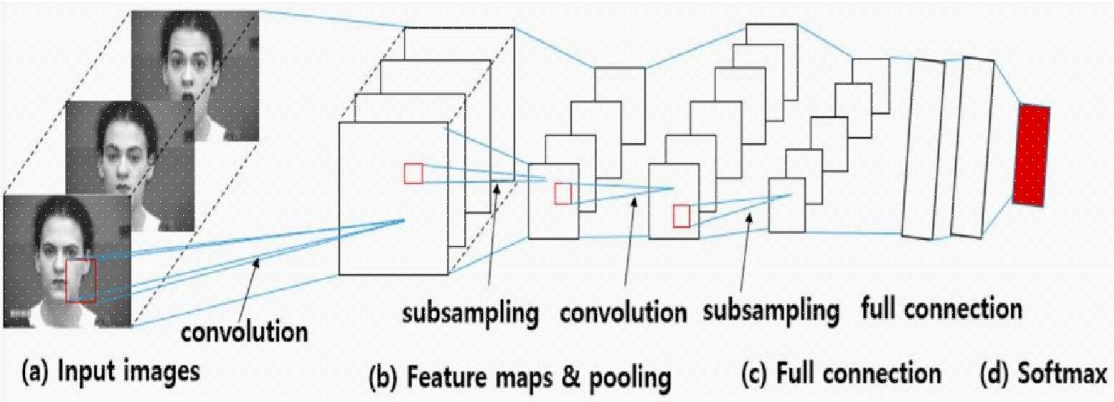

Fig-2 Procedure of CNN-Based FER approaches

CHAPTER-2

Literature Survey

Matthew S. Ratliff and Eric Patterson conducted an experiment to recognize the emotion using facial expressions with active appearance models. In their paper a framework for the classification of emotional states, based on still images of the face was done. The technique involved the creation of an active appearance model (AAM) that was trained on face images from a publicly available database to represent shapes and texture variation key to expression recognition. AAM used a set of feature classification scheme which identifies the six basic emotions. In this approach facial expression was analyzed by the means of the movement of facial muscles. Use FACS parameters to store those values. FACS was used as a framework for classification. The AAM has the ability to aid in initial face search algorithms and in extracting important information from both the shape and texture (wrinkles, etc.) of the face that may be useful for communicating emotion. So they adopted this technique as a feature extraction method. By giving the computer prior information such as eating habits, stress levels, sleep habits, etc. the ANN predict the emotional state of the user and can change its responses accordingly. In this approach we use the facial expression database known as “FEEDTUM.” This database contains still images and video sequences of eighteen test subjects, both

male and female, of varying age. Rather than hiring actors to artificially create or mimic emotional responses, this database was developed with the attempt to actually elicit the required emotions. Using a camera mounted on a computer screen, subjects show various movie clips that hopefully trigger an emotional response. The database is organized by category using the six basic emotions. Key areas were chosen to capture the movement of the brow, eyes, mouth, and nasal region as formed by the underlying muscles expected in expression of the face. Once an initial AAM was trained in several subjects, the search function helped automate the labeling process .

In the paper, Fully Automatic Recognition of Temporal Phases of Facial Actions, proposed by Michel F. Valstar and Maja Pantic the detection of a much larger range of facial behavior that include the classification of more facial expression other than the basic six emotions. The facial behavior is recognized by facial muscle actions which include action units (AUs).AUs help for higher decision making systems like emotion recognition system. Michel F. Valstar and Maja Pantic proposed fully automatic method that allows the recognition of 22 AUs and it also allows storing the temporal characteristics (which include the temporal segments like neutral, onset, offset and apex). For calculating the temporal features the proposed system uses a facial point detector which used Gabor-feature –based

booster classifier that automatically localize 20 facial fiducially points. The points that are detected from the face are tracked in an entire image sequence. To encode AUs and their temporal activation models, it uses a combination of support vector machines, Gentle Boost, and hidden Markov models. The five important steps that are used for the recognition of AUs and their temporal activation models are: registration and smoothening, midlevel parametric representation, facial AU classification, temporal activation models of facial AUs and finally emotion detection .

In the paper emotion recognition using facial expressions, speech and multimodal information recognition of human emotion based on facial expression or speech is done. Only limited system or work has been done to fuse this information. The accuracy and robustness of emotion recognition can be increased by combining these two techniques. This paper explains about the two approaches used to fuse the two modalities: decision level and feature level integration. The database holds the four emotions to classify: sadness, anger, happiness and neutral state. The database used in this technique is recorded by an actress. By the use of markers on the face, detailed facial motions can be captured with motion capture. The paper reveal that the system based on facial expression gave better performance than the system based on just acoustic information for the emotions considered.

The paper also shows the complimentary of the two modalities and that when these two modalities are combined, the robustness and the performance of the emotion recognition system improve measurably. The methodology used on the system based on speech includes the cues for audio emotion recognition and they are global level prosodic features such as the statistics of the pitch and the intensity. Therefore, the standard deviations, the ranges, the maximum values, the minimum values and the medians of the pitch and the energy were computed. In addition, the voiced/speech and unvoiced/speech ratio were also estimated.

By the use of sequential backward features selection technique, an 11-dimensional feature vector for each utterance was used as input in the audio emotion recognition system. In the case of a system based on facial expression the spatial data collected from markers in each frame of the video are reduced into a 4-dimensional feature vector per sentence, which is then used as input to the classifier. After the motion data are captured, they are normalized by the following steps:

(1) all markers are translated in order to make a nose marker be the local coordinate center of each frame,

(2) a frame with neutral and close-mouth head pose is picked as the reference frame,

(3) three rigid markers define a local coordinate origin for each frame, and

(4) each frame is rotated to align it with the reference frame. Each data frame is divided into five blocks: eyebrow, low eye, forehead, right cheek and left cheek area.

2.1 Comparative Study:

Different approaches which are followed for Facial Expression Recognition:

Neural Network Approach: The neural network contained a hidden layer with neurons. The approach is based on the assumption that a neutral face image corresponding to each image is available to the system. Each neural network is trained independently with the use of on-line back propagation.

Principal of Component Analysis: Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variable called Principal Components.

Support Vector Machine : In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Training & Testing Database: In machine learning, the study and construction of algorithms that can learn from and make predictions on data is a common task. Such algorithms work by making data driven predictions or decisions, through building a mathematical model from input data. The data used to build.

We will compare our research work with another research works and applications. We are actually study that research work and try to learn about them. We studied so many things from that research work and find many things such as algorithm, accuracy of that apps etc. all this research work are emotions analysis.

- "Infant facial expression and cries" in this paper author worked with infant emotions. Author work with infants eyes and mouth .author used clustering method, harmonic spectrum method. The accuracy of that Research is 75.2%.
- "Emotion recognition by dynamic HOG features" in this paper author elaborated an emotion recognition framework using dynamic dense grid-based HoG features. Proposed method performs better than static Uniform LBP implementation, used in the "baseline method" offered by the challenge organizers.

The accuracy of emotion detection is 70%.

- “Emotion based music player for android” in this paper author worked with emotion. Author work with viola and Jones algorithm, support vector machines (SVM) method.
- “Emotion recognition system for mobile application” in this paper author work with facial emotions. Author use viola and Jones algorithm .in this research author found 92.7% accuracy.
- “Automatic Recognition of Facial Displays of Unfelt Emotions” in this paper author work with emotions. Author was worked with simulation theory and also work with felt and unfelt emotions. After testing 51% accuracy was gain.
- “An Emotion Recognition Challenge” in this paper author used baseline method, principal component analysis (PCA) algorithm. The accuracy of thisresearch work is 62.3%.
- “Mood Prediction from Facial Video with Music “Therapy” on a Smartphone” This paper presents a prototype desktop version and a Smartphone app which analyses the mood of a video and predict the user’s mood. The app can play songs and change the song according to the mood analyzed. Accuracy of this work is 60%.

Image Processing is a useful method for performing different operations on an image to get a better image or to get some useful information from it. Normally image processing method consider an images as a two dimensional signals. Because of this usefulness of image processing, in our research we are dealing with this method. Mainly the project aim is to detect human's facial expression by applying image processing techniques and send them a massage about their internal feelings based on their facial expression. Those people who remain submerged in despair, this application is more beneficial for them. This application can get rid of their stress by playing music or jokes automatically. We hope that this application will bring a significant change of human life.

2.2 **EXISTING SY STEM:**

There have been constant improvements in image processing and emotion recognition techniques over the past years but yet some constraints still exist in the present systems which refrain from getting efficient images recognitions. Some of the constraints are:

- Low accuracy- Face recognition has low accuracy compared to

the proven performance of finger print and iris recognition.

- Variety of images of single faces- There are many attributes leading to the variability of images of a single face that add to the complexity of the recognition problem if they cannot be avoided by careful design of the capture situation.
- Personal changes- Change in facial expressions, aging and other personal factors also add to the difficulty in recognizing faces.

- Camera variations- Different cameras carry different lenses which again has a great impact on the picture being captured as different lens have different power. So any change in the camera or the lenses adds to the difficulty.

- FALSE ACCEPT- Impersonating somebody else is not a difficult task anymore. It can be easily achieved in today's time. This imitation if perfectly done by the attacker can easily con the system. This is called FALSE ACCEPT, Similarly many emotions and expressions can be wrongly classified sometimes.

- FALSE REJECT- A little expression or accessory can change the whole look of a person. This leads to unnecessary dismissal of the same face. This is called FALCE REJECT.

2.3 **PROPOSED SYSTEM:-**

There have been several advances in the past few years in terms of face detection, feature extraction mechanisms and the techniques used for expression classification, but development of an automated system that accomplishes this task is difficult. In this paper, we present an approach based on Convolutional Neural Networks (CNN) for facial expression recognition. The input into our system is an image; then, we use CNN to predict the facial expression label which should be one these labels: anger, happiness, fear, sadness, disgust and neutral.

My goal for this project is:

- Handle partial occlusions better.
- Make it more robust (lighting conditions etc.)
- More person independent
- Apply emotion recognition in applications. For example games, entertainment.

2.4 ARCHITECTURE DIAGRAM FOR PROPOSED METHOD:-

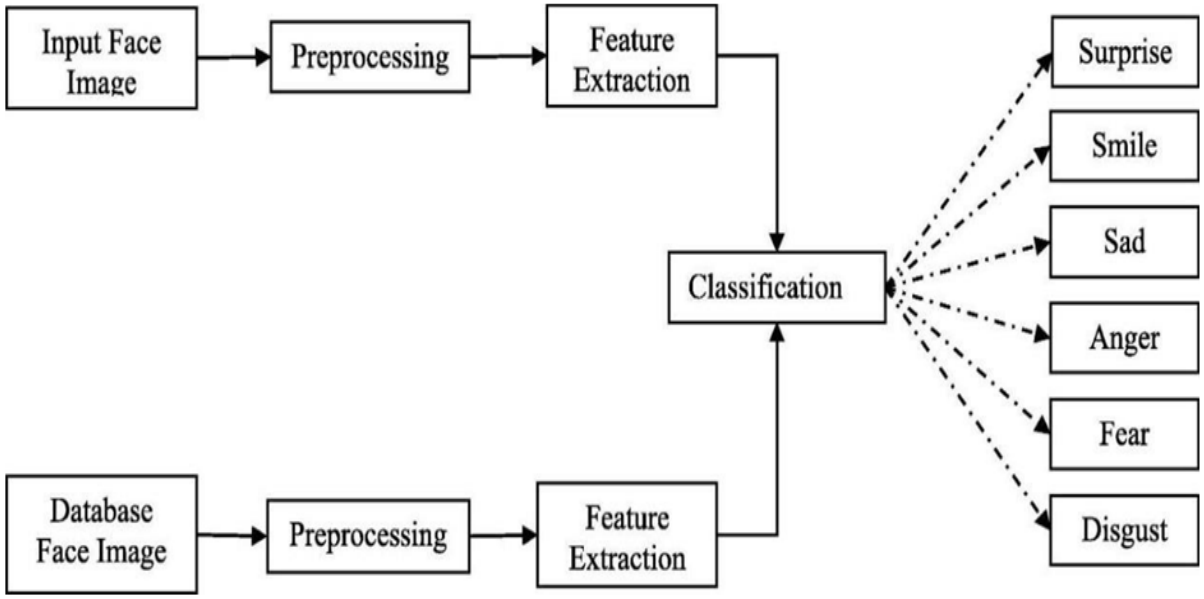

Fig-3 Architecture Diagram

CHAPTER-3

System Requirement

3.1 Introduction:-

Image processing is a technique than can convert an image into digital form and perform different kinds of operation on it for getting better image and useful information. Image processing technique used two types of method. These are analog and digital image processing. Analog technique can be used for hard copies and digital technique used for manipulating digital image. The purpose of image processing is divided into five groups. These are: visualization, image sharpening, image retrieval, measurement of pattern, image recognition. Visualization observe invisible object. Image sharpening that makes better image. Image retrieval finds interesting image. Measurement of pattern measures different objects of an image. Image recognition finds the difference of an image.

3.1.1 Machine Learning Algorithms:-

One of the most important applications of artificial intelligence is machine learning. It provides the application that can automatically

learn and improve from experience without being apparently programmed. The learning process starts with observations or data. Such as, we can assume a good decision based on direct experience or instruction. The basic aim is to allow the device without human interruption. Mostly machine learning algorithm is classified into two types. These are supervised and unsupervised learning.

a) **Supervised Machine Learning Algorithms:**

Supervised learning algorithms able to do different analysis with new data based on what it learned from the past and can also predict future event. The supervised learning algorithms create deduced function for predicting the starting analysis of known training data and output values. After some effective training the system makes a target for any new inputs. The system is able to compare its output with correct output and also find error for modification.

b) Unsupervised Machine Learning Algorithms:

Unsupervised machine learning algorithms are used for training unclassified and those data which are not leveled. Unsupervised

learning is able to describe a secret shape from unleveled data. This system can't provide proper output but it is able to take important decision from data set for describing secret shape from unleveled data.

c) Semi-supervised Machine Learning Algorithms:

Semi-supervised machine algorithms lies between supervised and unsupervised learning. For training they use both leveled and unleveled data but in this training data there are a small amount of leveled data and a huge amount of unleveled data. By using this method the systems are able to develop learning exactitude. Normally semi-supervised learning algorithms are used when leveled data need proficient and relevant resource for training.

3.1.2 **DEEP LEARNING**:-

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It

is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

Deep learning achieves recognition accuracy at higher levels than ever before. This helps consumer electronics meet user expectations, and it is crucial for safety- critical applications like driverless cars. Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images.

➔ **Benefits of Deep Learning:-**

- Has best-in-class performance on problems that significantly outperforms other solutions in multiple domains. This includes speech, language, vision, playing games like Go

etc. This isn't by a little bit, but by a significant amount.

- Reduces the need for feature engineering, one of the most time-consuming parts of machine learning practice.

- Is an architecture that can be adapted to new problems relatively easily e.g. Vision, time series, language etc., are using techniques like convolutional neural networks, recurrent neural networks, long short-term memory etc.

➔ **DISADVANTAGE:-**

- Requires a large amount of data — if you only have thousands of examples, deep learning is unlikely to outperform other approaches.

- Is extremely computationally expensive to train. The most complex models take weeks to train using hundreds of machines equipped with expensive GPUs.

- Do not have much in the way of strong theoretical foundation. This leads to the next disadvantage.

- Determining the topology/flavor/training method/hyper parameters for deep learning is a black art with no theory to guide you.

- What is learned is not easy to comprehend. Other classifiers (e.g. decision trees, logistic regression etc) make it much easier to understand what's going on.

3.1.3 CONVOLUTIONAL NEURAL NETWORK (CNN):-

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptions designed to require minimal pre-processing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially

overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing.

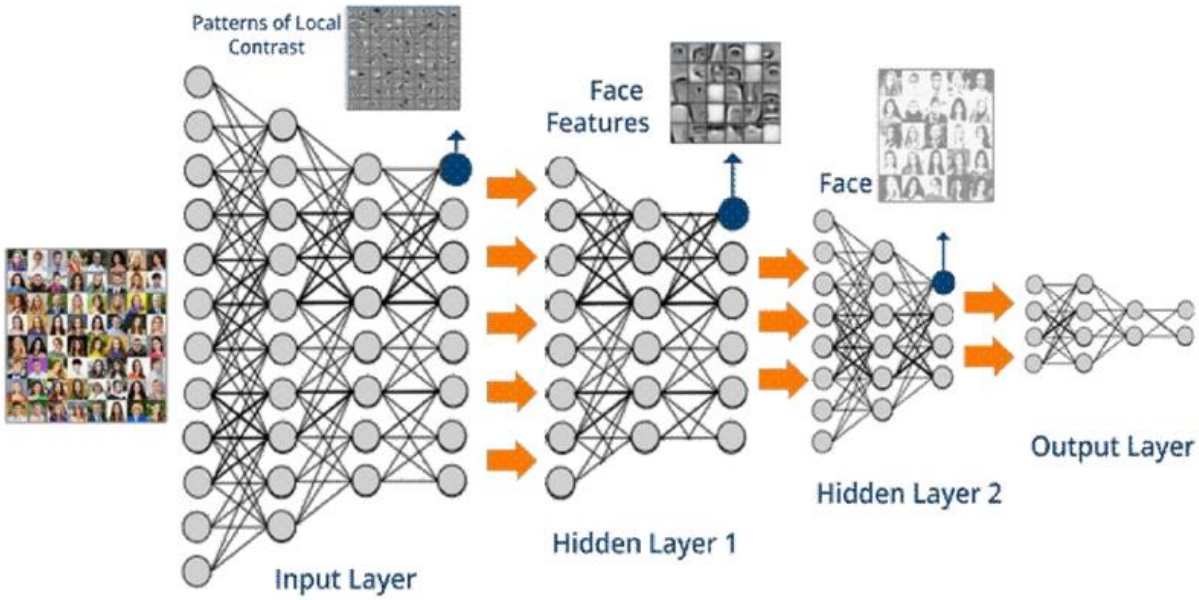

Fig-4 Structure of CNN

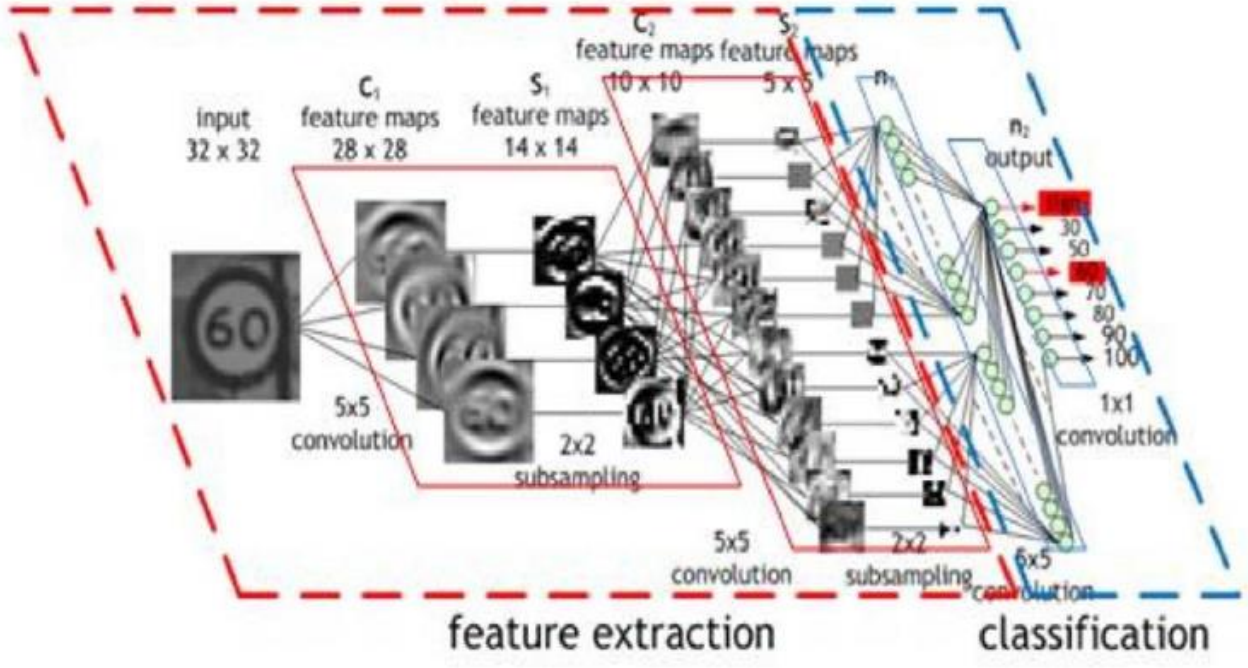

Fig-5 Example of CNN

3.2 PREREQUISITES OF THE PROJECT:

1. **Keras2.0.2:-**

   Keras is a high-level neural networks API, written in Python and capable of running on top of Tensor Flow, CNTK, or Theano .It was developed with a focus on enabling fast experimentation. Keras is an open source neural network library written in Python .It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or MXNet. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. It was developed as part of the research effort of project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), and its primary author and maintainer is François Chollet, a Google engineer.

2. **Tensorflow1.1.0:-**

TensorFlow is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from

the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.

3. **Pandas 0.19.1:-**

Pandas is an open source, BSD-licensed library providing high-performance, easy to-use data structures and data analysis tools for thePython programming language.

Pandas is aNumFOCUS sponsored project. pandas is an open source, BSD licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. pandas is a NumFOCUS sponsored project.

4. **Opencv2-python 3.2.0:-**

OpenCV-Python is a library of Python bindings designed to solve computer vision problems. Python is a general-purpose programming language started by Guido van Rossum that became very popular very quickly, mainly because of its simplicity and code readability. It enables the programmer to express ideas in fewer lines of code without reducing readability.

OpenCV-Python makes use of Numpy, which is a highly optimized library for numerical operations with a MATLAB-style syntax. All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib. OpenCV supports a wide variety of programming languages such as C++, Python, Java, etc., and is available on different platforms including Windows, Linux, OS X, Android, and iOS. Interfaces for high-speed GPU operations based on CUDA and OpenCL are also under active development. OpenCV-Python is the Python API for OpenCV, combining the best qualities of the OpenCV C++ API and the Python language.

5. **PyCharm3.5:-**

PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django. We have used PyCharm as the IDE in our project.

3.3FUNCTIONAL REOUIREMENTS:-

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases.

Here, the system has to perform the following tasks:

- Take the real time input of the person from the web cam.
- Identify the face and extract the facial features
- Based on the trained data, classify the emotion/gender and if any object seen.
- The recognized emotion is given as output in the form of a speech

3.4 NON-FUNCTIONAL REOUIREMENTS:-

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. This should be contrasted with functional requirements that define specific behavior or functions. The plan for implementing functional requirements is detailed in the system design. The plan for implementing non-functional requirements is detailed in the system architecture. Other terms for non-functional requirements are "constraints", "quality attributes", "quality goals", "quality of service requirements" and "non-behavioral requirements". Some of the quality attributes are as follows:

3.4.1 ACCESSIBILITY:-

Accessibility is a general term used to describe the degree to which a product, device, service, or environment is accessible by as many people as possible.

The system will be accessible to a lot of people as it will be incorporated with virtual assistants and home service robots.

3.4.2 MAINTAINABILITY:-

In software engineering, maintainability is the ease with which a software productcan be modified in order to:

- Correct defects
- Meet new requirements

Since the project will be implemented using python libraries it is easier to add or modifythe code since it won't have a larger code.

3.4.3 SCALABILITY:-

System is capable of handling increase total throughput under an increased load when resources (typically hardware) are added. System can work normally under situations such as low bandwidth and large number of users

3.4.4 PORTABILITY:-

Portability is one of the key concepts of high-level programming. Portability is the software code base feature to be able to reuse the existing code instead of creating new code when moving software from an environment to another. Project can be executed under different operation conditions provided it meet its minimum configurations. Only system files and dependant assemblies would have to be configured in such case.

3.5 HARDWARE REQUIREMENTS:-

Processor : Any Processor above 500 MHz

RAM : 512Mb

Hard Disk : 10 GB

Input device : Standard Keyboard & Mouse, Webcam

Output device : VGA and High Resolution Monitor

3.6 SOFTWARE REQUIREMENTS:-

Operating system : Windows XP or above

Front End : Python

IDE : Pycharm

CHAPTER-4

Project Design

The Steps Followed By the are as follows :

- Preprocessing : The image is loaded , it is then resized to maintain consistency , it is the converted to an numeric numpy array because DL models only understands numeric values.
- Input Layer: It contains raw pixel values of the image. Preprocessing is done before feeding pixels onto this layer.
- Convolution layer: This layer computes the dot product between the weights and a small region to which the neurons are connected on the input layer. It is followed by a pooling layer that samples down the dimensions along the width and the height to reduce the computational time due to a large number of convolution layers.
- Dense layer: It transforms the features through layers connected with trainable weights. This layer identifies the sophisticated features of the image that brings out the entire image.
- Output Layer : It is connected to previous layer and output the required classes or their probabilities.

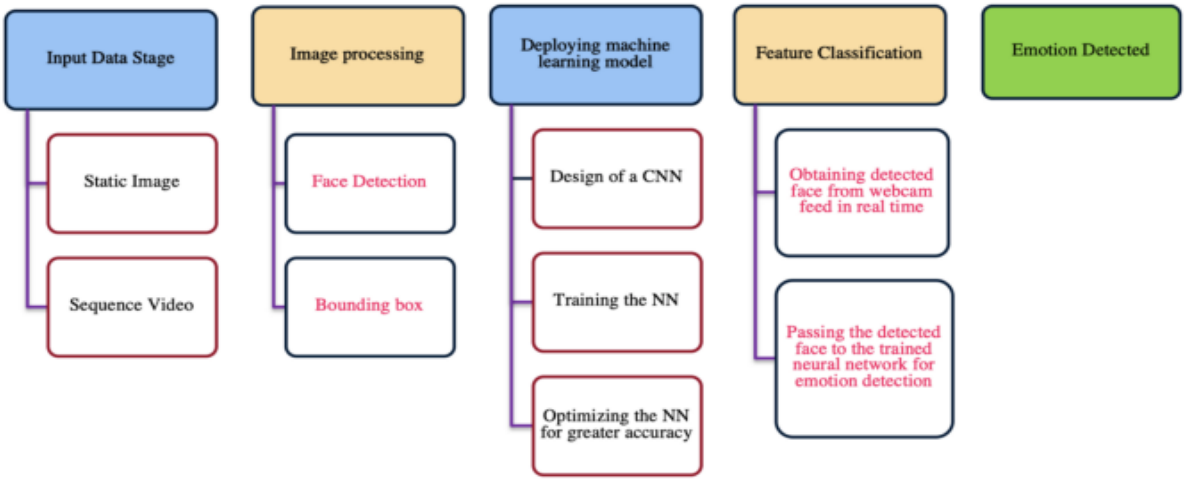

Fig 6: Project Design

- Preprocessing :

    1. Read the picture files (stored in data folder).
    2. Decode the JPEG content to RGB grids of pixels with channels.
    3. Convert these into floating-point tensors for input to neural nets.
    4. Rescale the pixel values (between 0 and 255) to the [0, 1] interval (as training neural networks with this range gets efficient).

- Feature Extraction :

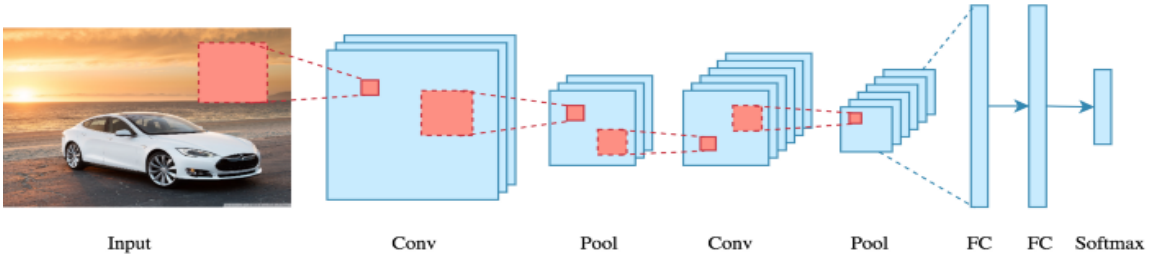

Fig 7 CNN Architecture

1. Convolution :Convolution is a mathematical operation to merge two sets of information. In our case the convolution is applied on the input data using a convolution filter to produce a feature map.
2. Non Linearity : For any kind of neural network to be powerful, it needs to contain non-linearity. We again pass the result of the convolution operation through relu activation function. So the values in the final feature maps are not actually the sums, but the relu function applied to them.
3. Stride And Padding : Stride specifies how much we move the convolution filter at each step. By default the value is 1.
4. Pooling : After a convolution operation we usually perform pooling to reduce the dimensionality. This enables us to reduce the number of parameters, which both shortens the training time and combats over fitting.
5. Fully Connected Layer : After the convolution + pooling layers we add a couple of fully connected layers to wrap up the CNN architecture. This is the same fully connected ANN.

- Classification : The Output Layer consisting of N(no of classes) neuron is present in the final layer. It uses soft max activation function which provides probability of each class such that it sums up to 1.

CHAPTER-5

Implementation /Working of Project

The system has achieved an overall accuracy of **68.85%** after training it for 20 epochs on a Windows system. The model can detect multiple faces in real time. Firstly, it identifies the emotion in each of the detected faces and then labels the emotion accordingly near the bounding box that appears around each detected face. The decision to build this on Python gives us the versatility to extend the project beyond mere facial emotion detection. We can build various applications with this project at the heart of each one of them.

→About Data set :

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Url : https://www.kaggle.com/msambare/fer2013

Source Code :

```
import pandas as pd

import numpy as np

import tensorflow.keras

import tensorflow

from tensorflow.keras.models import model_from_json

from tensorflow.keras.preprocessing import image

from deepface import DeepFace

from tensorflow.keras.callbacks import ModelCheckpoint

from tensorflow.keras.callbacks import TensorBoard

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Flatten, Dense, Dropout,BatchNormalization

from tensorflow.keras.layers import Convolution2D, MaxPooling2D, ZeroPadding2D,Conv2D,
Activation,MaxPool2D

from tensorflow.keras.utils import to_categorical

from tensorflow.keras.optimizers import SGD

import cv2


## Loading the File

emotion_data = pd.read_csv('/content/drive/MyDrive/Colab Notebook/SEM 7/fer2013.csv')

emotion_data.head()


## Image Preprocessing

X_train = []

y_train = []

X_test = []

y_test = []
```

```
for i in range(emotion_data.shape[0]):
    k = emotion_data['pixels'].iloc[i].split(" ")
    if emotion_data['Usage'].iloc[i] == 'Training':
        X_train.append(np.array(k))
        y_train.append(emotion_data['emotion'].iloc[i])
    elif emotion_data['Usage'].iloc[i] == 'PublicTest':
        X_test.append(np.array(k))
        y_test.append(emotion_data['emotion'].iloc[i])
X_train = np.array(X_train)
y_train = np.array(y_train)
X_test = np.array(X_test)
y_test = np.array(y_test)
X_train = X_train.astype('float32')
X_test = X_test.astype('float32')

X_train = X_train.reshape(X_train.shape[0], 48, 48, 1)
X_test = X_test.reshape(X_test.shape[0], 48, 48, 1)

y_train= to_categorical(y_train, num_classes=7)
y_test = to_categorical(y_test, num_classes=7)

## Model Creation
model = Sequential()
model.add(Conv2D(64, 5, data_format="channels_last", kernel_initializer="he_normal",
                 input_shape=(48, 48, 1)))
model.add(BatchNormalization())
```

```
model.add(Activation("relu"))

model.add(Conv2D(64, 4))
model.add(BatchNormalization())
model.add(Activation("relu"))
model.add(MaxPool2D(pool_size=(2,2), strides=2))
model.add(Dropout(0.5))

model.add(Conv2D(32, 3))
model.add(BatchNormalization())
model.add(Activation("relu"))

model.add(Conv2D(32, 3))
model.add(BatchNormalization())
model.add(Activation("relu"))
model.add(MaxPool2D(pool_size=(2,2), strides=2))
model.add(Dropout(0.5))

model.add(Flatten())

model.add(Dense(128))
model.add(BatchNormalization())
model.add(Activation("relu"))
model.add(Dropout(0.2))

model.add(Dense(7))
model.add(Activation("softmax"))
```

```
model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])
model.fit(X_train,
          y_train,
          epochs=20,batch_size = 100,validation_data = (X_test,y_test))
model.fit(X_train,
          y_train,
          epochs=40,batch_size = 100,validation_data = (X_test,y_test))
## Saving the Model
model = model_from_json(open("trainedModel.json", "r").read())
model.load_weights('model.h5')
## Loading the model for real time detection
face_haar_cascade = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')

cap = cv2.VideoCapture(0)

while True:
ret,frame = cap.read()
test_image = cv2.imread("happy face.jpg")
    result = DeepFace.analyze(frame,actions=['emotion'],enforce_detection=False)

converted_image = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

faces_detected = face_haar_cascade.detectMultiScale(converted_image)
    for (x, y, w, h) in faces_detected:
        cv2.rectangle(frame, (x, y), (x + w, y + h), (255, 0, 0))
```

```
roi_gray = converted_image[y:y + w, x:x + h]
roi_gray = cv2.resize(roi_gray, (48, 48))
image_pixels = tensorflow.keras.utils.img_to_array(roi_gray)
image_pixels = np.expand_dims(image_pixels, axis=0)
image_pixels /= 255

    predictions = model.predict(image_pixels)
max_index = np.argmax(predictions[0])
emotion_detection = ('angry', 'disgust', 'fear', 'happy', 'sad', 'surprise', 'neutral')
emotion_prediction = emotion_detection[max_index]

    ##cv2.putText(frame, emotion_prediction, (int(x), int(y)))

cv2.putText(frame,result['dominant_emotion'],(0,50),cv2.FONT_HERSHEY_SIMPLEX,2,(0,25
5,0),3)
resized_image = cv2.resize(frame, (1000, 700))
    cv2.imshow('Emotion', resized_image)
    if cv2.waitKey(10) == ord('b'):
        break
cap.release()
cv2.destroyAllWindows()
```

CHAPTER-6

Result or Output

```
epochs=20,batch_size = 100,validation_data = (X_test,y_test))

Epoch 1/20
288/288 [==============================] - 18s 63ms/step - loss: 0.9854 - accuracy: 0.6319 - val_loss: 1.1040 - val_accuracy: 0.5876
Epoch 2/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9729 - accuracy: 0.6342 - val_loss: 1.1016 - val_accuracy: 0.5904
Epoch 3/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9638 - accuracy: 0.6379 - val_loss: 1.0803 - val_accuracy: 0.5974
Epoch 4/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9569 - accuracy: 0.6395 - val_loss: 1.0824 - val_accuracy: 0.5965
Epoch 5/20
288/288 [==============================] - 18s 63ms/step - loss: 0.9408 - accuracy: 0.6472 - val_loss: 1.0724 - val_accuracy: 0.6030
Epoch 6/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9306 - accuracy: 0.6492 - val_loss: 1.0886 - val_accuracy: 0.5926
Epoch 7/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9315 - accuracy: 0.6509 - val_loss: 1.0693 - val_accuracy: 0.6010
Epoch 8/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9141 - accuracy: 0.6561 - val_loss: 1.0766 - val_accuracy: 0.6055
Epoch 9/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9078 - accuracy: 0.6594 - val_loss: 1.1664 - val_accuracy: 0.5765
Epoch 10/20
288/288 [==============================] - 18s 62ms/step - loss: 0.9009 - accuracy: 0.6593 - val_loss: 1.0612 - val_accuracy: 0.6043
Epoch 11/20
288/288 [==============================] - 18s 63ms/step - loss: 0.9004 - accuracy: 0.6606 - val_loss: 1.0652 - val_accuracy: 0.5999
Epoch 12/20
288/288 [==============================] - 18s 63ms/step - loss: 0.8813 - accuracy: 0.6706 - val_loss: 1.1007 - val_accuracy: 0.5932
Epoch 13/20
288/288 [==============================] - 18s 63ms/step - loss: 0.8837 - accuracy: 0.6682 - val_loss: 1.0469 - val_accuracy: 0.6121
Epoch 14/20
288/288 [==============================] - 18s 63ms/step - loss: 0.8753 - accuracy: 0.6730 - val_loss: 1.0641 - val_accuracy: 0.6030
Epoch 15/20
288/288 [==============================] - 18s 63ms/step - loss: 0.8683 - accuracy: 0.6738 - val_loss: 1.0520 - val_accuracy: 0.6063
Epoch 16/20
288/288 [==============================] - 18s 62ms/step - loss: 0.8575 - accuracy: 0.6775 - val_loss: 1.0646 - val_accuracy: 0.6069
Epoch 17/20
288/288 [==============================] - 18s 62ms/step - loss: 0.8439 - accuracy: 0.6829 - val_loss: 1.0506 - val_accuracy: 0.6163
Epoch 18/20
288/288 [==============================] - 18s 62ms/step - loss: 0.8412 - accuracy: 0.6841 - val_loss: 1.0909 - val_accuracy: 0.5979
Epoch 19/20
288/288 [==============================] - 18s 62ms/step - loss: 0.8411 - accuracy: 0.6856 - val_loss: 1.0614 - val_accuracy: 0.6133
Epoch 20/20
288/288 [==============================] - 18s 62ms/step - loss: 0.8331 - accuracy: 0.6885 - val_loss: 1.0614 - val_accuracy: 0.6169
<keras.callbacks.History at 0x7f99b53712d0>
```

Fig 8: Accuracy achieved

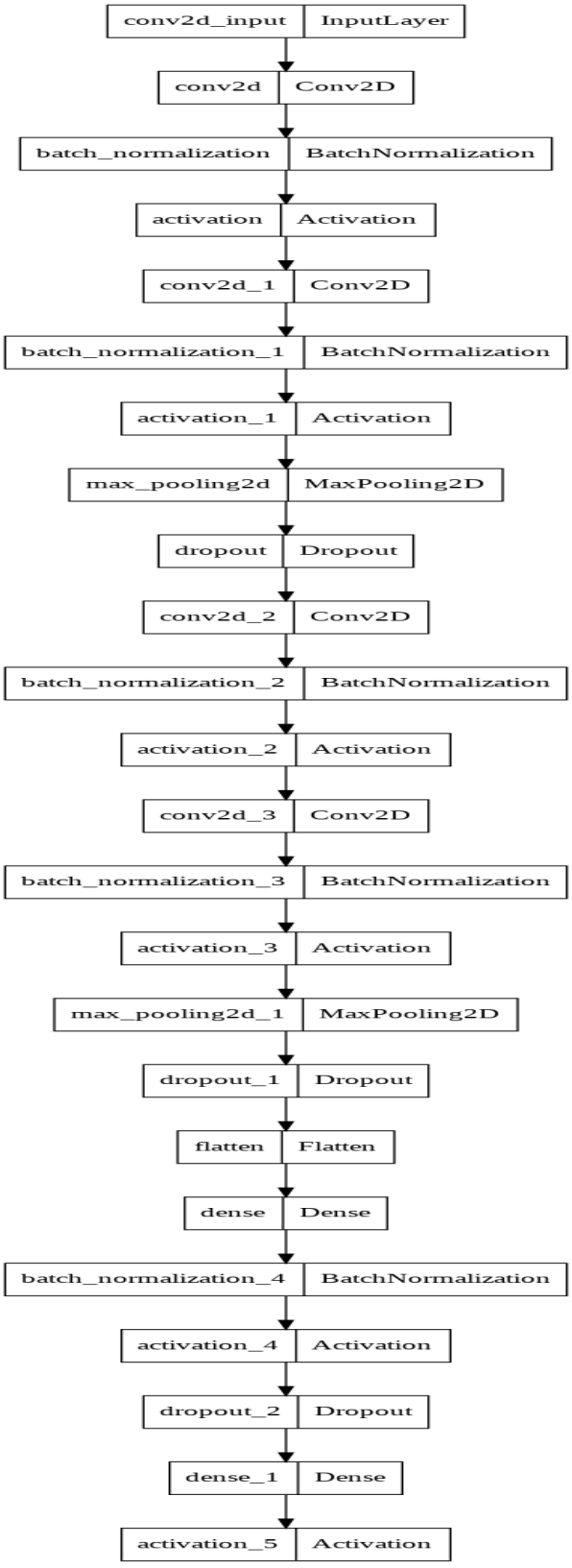

Fig 9: Model diagram

```
[18]  cm = confusion_matrix(test,pred)

      cm

      array([[239,   3,  48,  33,  89,  10,  45],
             [ 21,  19,   4,   2,   4,   1,   5],
             [ 65,   3, 166,  30, 131,  41,  60],
             [ 35,   0,  31, 719,  38,  15,  57],
             [ 84,   1,  78,  31, 345,  12, 102],
             [ 20,   0,  44,  28,  16, 297,  10],
             [ 54,   1,  44,  52, 126,   8, 322]])
```

Fig 10: Confusion Matrix

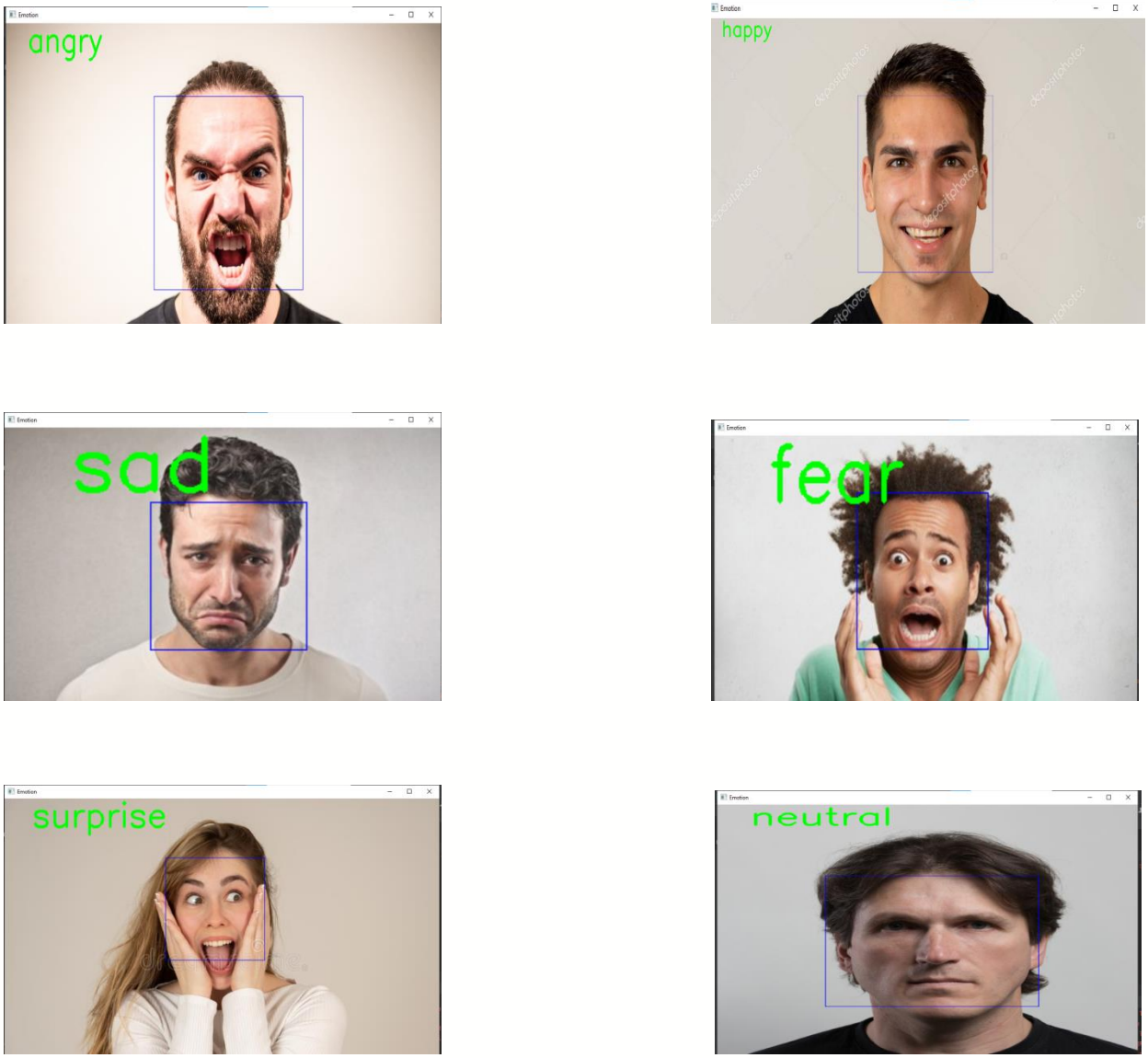

Fig 11: Sample Outputs

Chapter 7

Conclusion

In this project, we have proposed a deep learning based facial emotion detection method from image. We discuss our proposed model using dataset FERC-2013. The performance evaluation of the proposed facial emotion detection model is carried out in terms of validation accuracy, computational complexity, detection rate, learning rate, validation loss, computational time per step. We analyzed our proposed model using trained and test sample images, and evaluate their performance compare to previous existing model. Results of the experiment show that the model proposed is better in terms of the results of emotion detection to previous models reported in the literature. The experiments show that the proposed model is producing state-of-the-art effects on the dataset.

In this case, when the model predicts incorrectly, the correct label is often the second most likely emotion. The facial expression recognition system presented in this research work contributes a resilient face recognition model based on the mapping of behavioural characteristics with the physiological biometric characteristics.

The physiological characteristics of the human face with relevance to various expressions such as happiness, sadness, fear, anger, surprise and disgust are

associated with geometrical structures which restored as base matching template for the recognition system. The behavioural aspect of this system relates the attitude behind different expressions as property base. The property bases are alienated as exposed and hidden category in genetic algorithmic genes.

The gene training set evaluates the expressional uniqueness of individual faces and provide a resilient expressional recognition model in the field of biometric security. The design of a novel asymmetric cryptosystem based on biometrics having features like hierarchical group security eliminates the use of passwords and smart cards as opposed to earlier cryptosystems. It requires a special hardware support like all other biometrics system. This research work promises a new direction of research in the field of asymmetric biometric cryptosystems which is highly desirable in order to get rid of passwords and smart cards completely. Experimental analysis and study show that the hierarchical security structures are effective in geometric shape identification for physiological traits.

The human face is an important organ of an individual‘s body and it especially plays an important role in extraction of an individual‘s behavior and emotional state. Manually segregating the list of songs and generating an appropriate playlist based on an individual‘s emotional features is a very tedious, time consuming, labor intensive and upheld task. We are using 3 types of algorithm in our development. Such as PCA, MPCA, Machine learning language. We are working

with human’s eyes and mouth for emotion detection. We are working and testing many images to detect human’s emotions. The accuracy of our research work is 80%. The human face is an important organ of an individual‘s body and it especially plays an important role in extraction of an individual‘s behavior and emotional state. Manually segregating the list of songs and generating an appropriate playlist based on an individual‘s emotional features is a very tedious, time consuming, labor intensive and upheld task. We are using 3 types of algorithm in our development. Such as PCA, MPCA, Machine learning language. We are working with human’s eyes and mouth for emotion detection. We are working and testing many images to detect human’s emotions. The accuracy of our research work is 80%.

CHAPTER-8

Future Scope

It is important to note that there is no specific formula to build a neural network that would guarantee to work well. Different problems would require different network architecture and a lot of trail and errors to produce desirable validation accuracy. This is the reason why neural nets are often perceived as "black box algorithms". In this project we got an accuracy of almost 70% which is not bad at all comparing all the previous models. But we need to improve in specific areas like-

- number and configuration of convolutional layers
- number and configuration of dense layers
- dropout percentage in dense layers

But due to lack of highly configured system we could not go deeper into dense neural network as the system gets very slow and we will try to improve in these areas in future. We would also like to train more databases into the system to make the model more and more accurate but again resources becomes a hindrance in the path and we also need to improve in several areas in future to resolve the errors and improve the accuracy.

The future scope in the system would to design a mechanism that would be automatic playing music or videos based on the human facial mood. This system would be also helpful in music therapy treatment and provide the music therapist the help needed to treat the patients suffering from disorders like mental stress, anxiety, acute depression and trauma.

CHAPTER-9

REFERENCES

[1] P. Pal, A. N. Iyer and R. E. Yantorno, "Emotion Detection From Infant Facial Expressions And Cries," *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings,* Toulouse, 2006,pp. II-II.

[2] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," *Face and Gesture 2011*, Santa Barbara, CA, 2011, pp.884-888.

[3] K. S. Nathan, M. Arun and M. S. Kannan, "EMOSIC — An emotion based music player for Android," *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Bilbao, 2017,pp. 371-276.

[4] M. S. Hossain and G. Muhammad, "An Emotion Recognition System for Mobile Applications,"in *IEEE Access*, vol. 5, pp.2281-2287, 2017.

[5] K. Kulkarni *et al*., "Automatic Recognition of Facial Displays of Unfelt Emotions," in *IEEE Transactions of Affective Computing*.

[6] Y. Li, J. Tao, B. Schuller, S. Shan, D. Jiang and J. Jia, "MEC 2017: Multimodal Emotion Recognition Challenge," *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, Beijing, 2018, pp.1-5

[7] Y. Seanglidet, B. S. Lee and C. K. Yeo, "Mood prediction from facial video with music “therapy” on a smartphone," *2016 Wireless Telecommunications Symposium (WTS)*, London, 2016, pp. 1-5.

[8] A. V. Iyer, V. Pasad, S. R. Sankhe and K. Prajapati, "Emotion based mood enhancing music recommendation," *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2017, pp. 1573-1577.

[9] T. Moriyama, K. Abdelaziz and N. Shimomura, "Face analysis of aggressive moods in automobile driving using mutual subspace method," *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, 2012, pp. 2898-2901

[10] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, Dec. 2000.

➔ **URL:**

- http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=895976&isnumber=19391
- https://ieeexplore.ieee.org/Xplore/home.jsp
- https://en.wikipedia.org/wiki/Main_Page
- https://www.google.com.bd/?gws_rd=cr&ei=sK0XWLPAEcvmvgT12o7AAg
- Etc……..