EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

A PROJECT REPORT

Submitted by

G. ADERSH [Reg No: RA1511008010047]
K.J. AJAY ESWAR [Reg No: RA1511008010049]
U. AJIT SUNDARESH [Reg No: RA1511008010055]
S. HARIHARAN SUBRAMANIAN [Reg No: RA1511003010021]

Under the guidance of

Dr. A. RAZIA SULTHANA

(Associate Professor, Department of Information Technology)

In partial fulfilment of the Requirements for the Degree

of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY

FACULTY OF ENGINEERING AND TECHNOLOGY

SRMINSTITUTE OF SCIENCE AND TECHNOLOGY

MAY 2019

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR-603203

BONAFIDE CERTIFICATE

Certified that this project report **"EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS"** is the bonafide work of **U. AJIT SUNDARESH (RA1511008010055), G. ADERSH (RA1511008010047), K.J. AJAY ESWAR (RA1511008010049), S. HARIHARAN SUBRAMANIAN (RA1511008010021),** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. A. RAZIA SULTHANA
GUIDE
Assistant Professor
Dept. of Information Technology

Dr. G. VADIVU
HEAD OF THE DEPARTMENT
Dept. of Information Technology

Signature of the Internal Examiner

Signature of the External Examiner

ABSTRACT

Human emotions are not consistent and they are actually a result of internal and external circumstances happening around an individual. Intense research and investment have been made on human emotions which can lead to a variety of applications. The existing system comprises of automatically creating a music playlist based on the genres, artists etc. Yet another option is manually segregating music files into playlists. Recent issues involve multiple frequency estimation and music similarity computation. A QBSH (Query by singing and humming) system determines a song by using its contents (tune and rhythm). But the issue revolving around this option is that they are time-consuming and not always satisfies the user. In the existing system, the emotion of the user is not taken into consideration. Since emotions play a vital role in day-to-day activities, a music recommendation system can be developed which takes the human emotions into account. By identifying the emotion of an individual, suitable music can be recommended. The system aims at examining the data provided by identifying the emotion of the user. A deep-learning algorithm is applied to classify the various emotions after which labels are generated and suitable music is played. The proposed system has delivered results with significant accuracy and it also paves the way for further research in this area.

ACKNOWLEDGEMENT

The completion of the project requires constant assistance and guidance from teachers and we are extremely happy to of received it from our project guide Dr. A. RAZIA SULTHANA, Assistant Professor, SRM Institute of Science and Technology. We owe the proper completion of the project to her valuable inputs and we would never forget all the right questions that she kept on asking us, the clues which she kept giving us which helped us on the long run in completing the project. We could not have imagined a better guide who can motivate us and advise us for our major project.

This Major Project has been a huge team-building exercise and we can't thank enough our faculties for helping us through the process. Even when the team-mates were out doing internship in other cities, our guide was very understanding and she took time off her busy schedule to read and reply to our mails or messages.

We also would like to show our gratitude to our Head of the Department, Dr. G. VADIVU, Department of Information Technology, SRM Institute of Science and Technology, for providing us with the right environment and the infrastructure to complete our project.

We would also like to appreciate the dates that were assigned for the project reviews and the project report submission. These dates were told in advance and they helped us in implementing an effective project plan to go about.

Author

TABLE OF CONTENTS

LIST OF TABLES

NIL

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS

1	QBSH	Query by singing and humming
2	ANN	Artificial Neural Networks
3	AAMM	Active Appearance Model Method
4	ESTM	Emotion State Transition Model
5	COMUS	Content based Music Recommendation
6	NVM	Non-Negative matrix Factorization
7	SVM	Support Vector Machine
8	EDM	Electronic Dance Music
9	VA	Valence Arousal
10	PDF	Probability Density Function
11	GSR	Galvanic Skin Response
12	PPG	Photo-Plethysmography
13	CNN	Convolutional Neural Network

CHAPTER 1

INTRODUCTION

Communication is essential for sharing of messages or resource among individuals. Information can be shared among individuals either verbally or nonverbally. Facial Expression of an individual can be very useful in tapping the subject's mood and behaviour. Human Emotion plays a crucial role in expressing the thought of an individual. They fall into one of the six basic types emotions which are sadness, happiness, anger, fear, disgust and surprise. These emotions can be determined by detecting changes in shape, size and movement of eyebrows, eyes and mouth. Modern devices that have access to internet contain millions of songs at any time. Music at any instant has the power to change the mood of an individual. Our main focus is to generate a music playlist which is automatically created by interpreting these human emotions.

Modern day technology includes automatic classification of a music based on Genres, artist, country, frequency and more. Mobile applications like Spotify, Saavn, wynk etc are some of the few applications who does automatic music classification. Here the users can create their own playlists or listen to the playlist generated by the application. Another modern-day application includes finding a song which a user is not able to identify. Mobile application like Shazam, trackID, Sound Hound can be very useful in such cases. Other Application like Musixmatch, YouTube music helps in displaying the lyrics of song at any instant and automatic translation of the lyrics from one language to another language. Music is said to be the greatest healing tool in the world. Therefore, we tap in that aspect of music to make wonders for the individual who is going through a certain emotion.

1.1 OVERVIEW OF NEURAL NETWORKS

A Human Body can be said to be one of the most important and complex system in the world. The complexity arises due to the innumerable amount of nerves that run through our body which carries and processes information. This also makes the human body to be one of the most intelligent species on Earth. A Neural Network can be said to be inspired from this biological information processing system. Research in this area has led to developing wide range of applications. So, just like how people learn from experience, a neural network also learns through experience i.e., training.

1.2 WHY NEURAL NETWORKS

Neural systems can be utilized to extricate designs and distinguish patterns that are too intricate to be in any way seen by either people or other machine techniques; with their exceptional capacity to get importance from convoluted or uncertain information.A prepared neural system can be considered as a specialist in the data class that was given for analysis.

Some of the main advantages are:

1. It has the ability to do lot of processes or operations with the help of knowledge gained from its past outcomes.
2. It has the ability to represent its own processed data or represent the data that is received at the time of processing.
3. It can perform various operations or tasks in parallel and this specific advantage can be incorporated into multiple systems which are designed for taking advantage of this capability.
4. Partial collapse of a network results in performance degradation. However, even with major architectural damage, some network capabilities may be kept as it is.

1.3 CLASSIFICATION

All grouping assignments rely upon marked datasets; that is, all together for a neural system to get familiar with the relationship amongst label and information, people need to exchange their insight to the dataset. It is also known as the Supervised Learning.

- Face detection, identifying all the individual's faces in images, Detecting the mood of the individual
- Object identification in images
- Gesture recognition in images/videos
- Sentiment recognition, voice detection
- Classification of spam in emails, fraudulent (in claims of insurance)

1.4 CLUSTERING

Clustering or grouping can be referred to as the fundamental data analysis method which detects similarities or finds pattern in the given data. This enables a system to learn without the use of labels. This type of learning is referred to as the unsupervised learning. And one of the most common unsupervised learning is the cluster analysis. Highly accurate models can be developed using this method of analysis.

- Search: Document comparison, finding similarity in images or sounds
- Anomaly detection: By looking for similarities, we can also detect anomalies i.e., unusual behaviour. By detecting such unusual behaviours, it helps us to prevent various frauds and also results in error detection/correction.

1.5 ELEMENTS OF NEURAL NETWORK

Neural network is a network that is composed of several layers. The name that is used to refer "stacked neural networks" is Deep Learning.

Each layer consists of multiple nodes which is responsible for various processes or computations. The concept behind these nodes are loosely based on the neurons present in the human brain. These neurons act or react based on the changes in behaviour i.e., a change in the stimuli. Various weights and coefficients are added to the input data present in each and every node. This can lead to the amplification or dampening of the input thereby adding significance to it. Later, this significance can help in regard to the operation that the algorithm is trying to learn. The products of the data input and the weights are calculated and then summed and this result is passed through the activation function which is present in the node. It also decides if it should and to what extent the data is to travel through the network in order to affect the final result. The neuron is said to be activated if the signal passes through it.

The Node Diagram may look like:

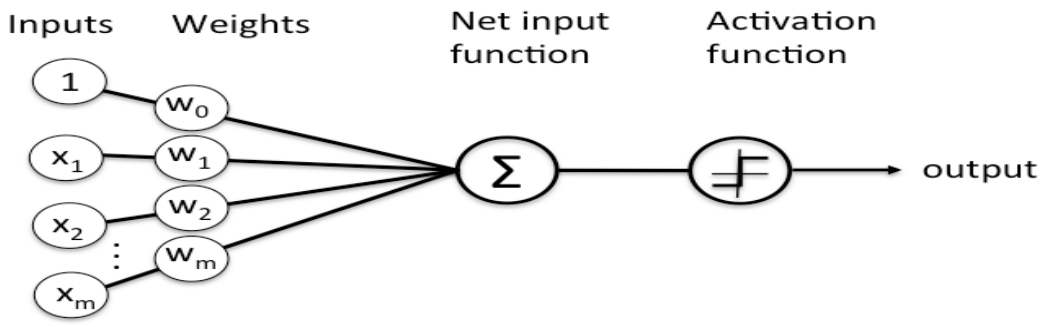

Fig 1.1 Node diagram of a neural-based network

Every layer in the node contain neuron-like switches which turn on or off as the signal passes through the network. The output that is generated at each layer becomes the input for the next layer.

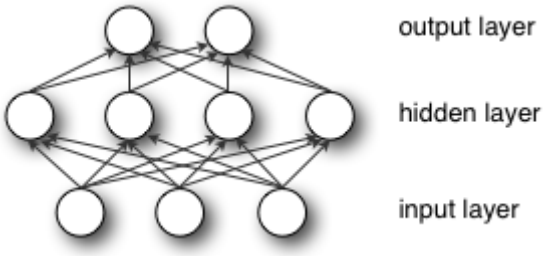

Fig 1.2 Different layers in a Neural-based Network

Thus, the classification in a neural network and input clustering to detect patterns and anomalies is done by assigning significance which is computed by pairing the input features with the weights of the model.

1.6 KEY CONCEPTS OF NEURAL NETWORKS

The number of node layers through which a data passes in a network is referred to as its depth. This feature is used to distinguish the deep learning networks from the single hidden layer networks. Depth of a neural network is very significant for pattern recognition. Early version of a neural network model was not very deep. It mostly composed just an input, output layer and a hidden layer sandwiched between them. But to qualify something as a deep learning network, it should at least consist of more than three layers (including input and output).

Every later in a deep learning network architecture is trained by recombining the output features which are received from its previous layer. Furthermore if we move into the neural system, more difficult highlights are perceived by neurons as they recombine highlights from the earlier layer.

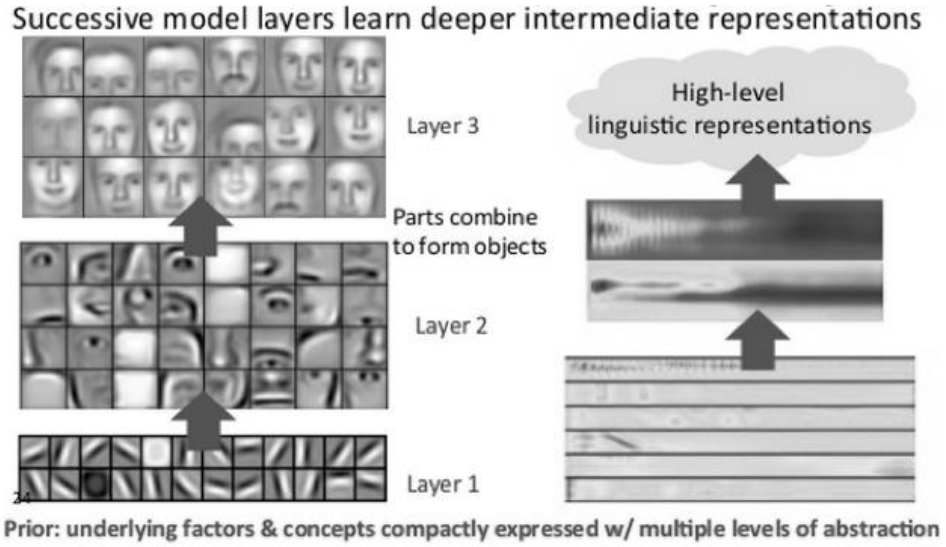

Fig 1.3 Feature Hierarchy in Image Processing

The above diagram represents the feature hierarchy. It is formed by the increasing level of complexity and abstraction. This is one unique feature of the deep-learning architecture which makes it handle large amounts of multi-dimensional data with billions of parameters.

Automatic feature extraction is possible in deep-learning networks. This characteristic makes it different from many machine-learning algorithms which require human involvement to some extent. So, this way, we can develop many smart systems. In a deep-learning network, features are trained and learnt automatically with the help of recombining the input from the previous samples. With this process they are also able to draw connections, recognize patterns and provide optimal results.

The output layer in a deep-learning network consists of a particular outcome or label. That outcome is used in prediction, for instance, given an image and the output we receive, we can declare that the given input is 85% likely to represent a person. In case of emotion analysis based on an image, the output may consist of a probability value which is achieved by using SoftMax layer in the architecture.

1.7 EMOTION ANALYSIS

Emotions play a major role in how an individual lets his feelings known to other people. It also affects how they live and communicate with others. One can say that we are all ruled by various sorts of these emotions. These emotions determine how we react to every action and they maybe both voluntary and involuntary depending upon the action. Every decision we make, every action we perform are all determined by the emotions which we are going through in that moment. It's also why they say never take a decision when you are angry.

1.8 TYPES OF EMOTIONS

It's been a research area for many Psychologists to identify how, when and why people emote a particular way to every situation. They have compiled various types of emotions that people undergo. They have also classified these emotions to better explain how humans feel.

Basic Emotions: Paul Eckman, a popular psychologist in the 1970s classified six emotions that he thought were universally experienced by all human creatures irrespective to the various cultures around the globe. The emotions which he classified were sadness, anger, happiness, disgust, anger and fear. He proposed that be it any kind of situation, any human will react with one of these emotions. He concluded that these emotions were appealed to all the human beings.

Combining Emotions: Robert Plutchik, another famous psychologist proposed something called as the "wheel of emotions". He said that sometimes these emotions can be combined to form different emotions. He compared it with colours that when they are mixed together with another colour, they give rise to a new colour. Similarly, when emotions mix together, they give rise to new emotions. These basic emotions act as building blocks. For example, anger and disgust can be combined to create hate.

1.9 OVERVIEW ON EMOTION ANALYSIS

The emotional analysis is an integral part of Affective Computing. We may commonly know it as artificial emotional intelligence or emotion AI. It involves gathering data from faces, body language or voices to identify the human emotion. It is an area of study within cognitive computing. Emotion analysis is performed by a range of human-computer interactions which enables a computer to identify and create a response appropriate to the user.

This is where Deep-learning algorithms come into play which is used to train a model that can detect various emotions. Psychologically, the word 'affect' is used to refer to the emotional tone of a patient. This area of emotional analysis has been very important to perform many researches and come up with wide range of applications.

One such application is the 'Emotion based Music Recommendation System', a prototype which we have developed by performing emotion analysis using Convolutional Neural Networks.

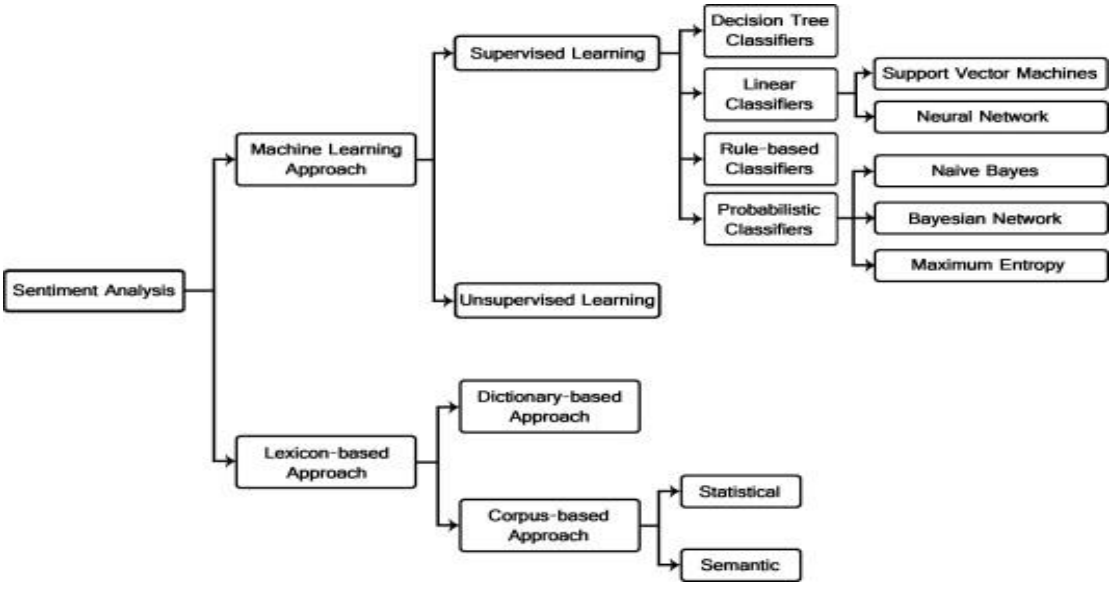

Fig 1.4 Different ways of performing sentiment analysis

CHAPTER 2

LITERATURE SURVEY

2.1 PREVIOUS WORK RELATED TO PROPOSED SYSTEM

1. For Automatic Facial Expression recognition this research paper [1] uses three phases. These three phases are 1. Face detection2. Feature Extraction and 3. Expression recognition. In the First Phase, YCbCr Colour model are used for face detection, lighting compensation for obtaining face and morphological operations for holding required features of the face i.e. eyes, eyebrows and mouth. This System also uses Active Appearance Model Method (AAM) for facial feature extraction. In this method the features on the face like eye, eyebrows and mouth are located and a data file is created which gives information about the model points detected. Different facial expressions are given as input to the AAM Model which changes according to expression.

2. Three different ways are used in this paper [2] for emotion classification and context-based music recommendation. They are 1. EmotionStateTransitionModel (ESTM) 2. Context-based music recommendation (COMUS) 3. Nonnegative matrix factorization (NMF). ESTM is predominantly used to model various human emotions and their transition to music. It acts like a bridge between an individual's mood and low-level music features. With the help of ESTM the most legitimate music can be recommended to the client for travelling to the ideal state.COMUS ontology is utilized for demonstrating user's musical inclinations and setting, and for supporting thinking about the client's ideal feeling and inclinations. COMUS are a music dedicated ontology developed by including particular classes for music suggestion which incorporates mood, situation and other features. In order to reduce the dimensions data's related to music are gathered after which NMF are applied to map them to ESTM

3. The main objective of this paper [3] is to generate a music recommendation system by observing the sentiments of user and polarity of words used in social media. Sentiment intensity metric (Sentimeter-Br2) is used to extract an individual emotion from SocialNetworks.Sentiment-Br2 is a sentiment intensity metric whose main goal is to improve the overall accuracy and efficiency of music recommendation system. The words extracted from social media is ranked positive, negative or neutral based on sentiment intensity, according to which a musical playlist is generated and played to the respective user.

   A framework is created where the user registers by giving the necessary details creates a login account. Every time the user posts some content the phrases used by him are collected and stored. These words are then analysed on a day to day basis and are classified by the sentiment metric system. Based on the mood of the individual and his/her preference a playlist is generated and played to the user. The results showed that 72.5% of the total number of users considered the proposed recommendation system to be useful than the traditional recommendation system.

4. The goal of this research paper [4] is torecommend songs that the user likes, songs which are fresh or new to the user's ear and fit the user's listening pattern. The system mainly focuses on behaviour of the user and metadata rather than the content. A Forgetting Curve is used to estimate the freshness of a song and evaluate likeness using user log. The user's behaviour on the song is continuously monitored which is used as feedback for suggesting better songs when the user is not in a good mood or not satisfied. If a user listens to song completely it means that the user likes that song and similar kinds of songs are recommended to the user.

   On the other hand, if the user skips a song the system infers that the user dislikes that song and is less recommended. Thus, the user's attitude towards a song is evaluated continuously on a long-term basis. The five factors which are important in designing recommendation systems are freshness, year, favour, time pattern and genre. Lesser the feedback better the automatic music recommendation system.Results showthat the recommendation system surpasses the baseline and is proven to be effective.

5. This research paper [5] uses real time datasets for music recommendation system. A TV music program's audience were requested to rate the music of the participants based on their music preference and emotional feelings. The developers targeted low-level music features which triggered human emotions among the audience. In addition to this a personalized music recommendation system was implemented using low-level music features, listener's history and content analysis.

   For selection of low-level music features which are responsible for triggering emotions, preference analysis method is used which is based on empirical evaluation scores. Once the features are selected a design is created based on the selected features, listening history which are combined with environmental information. The design shows the subjective validity and accuracy of audience evaluation. Though aural aspect affected a large part of the evaluation, by extracting more features and increasing the size of dataset better results with high accuracy can be obtained.

6. All the existing fields in this world have entered the digital era. Music too has a new genre called EDM (Electronic Dance Music) where the music is composed with the help of software. With the evolution of digital music, music recommendation system will be highly useful for the users. This paper [6] aims at developing a novel model for emotion-based music recommendation by association discovery from film music. Content based filtering and collaborative filtering are the two approaches considered for music recommendation. Content based monitors the songs the user listens to on a day basis and recommends music according to that. On the other hand, collaborative filtering recommends music to peer groups based on their preference. By modifying the affinity graph and examining the relationship between music features and emotions from film music results show that the proposed design achieves 85% accuracy.

7. Due to the wide applications of music retrieval and recommendation, computation modelling has been extensively studied in recent years. Even though significant progress has been made, due to the difficulty in detecting an emotion this task remains highly challenging. Since emotions perceived by users are subjective by nature it makes the process of tapping the emotion of the user and implementing the predictive model more difficult.

   In this paper [7] a novel machine learning approach is proposed that differentiates the music emotion as a probability distribution function in the valence-arousal emotion space. Emotion from a music piece is represented as a probability density function (PDF) in the valence arousal space from human annotations. By optimizing objective functions of music pieces combination coefficients are studied to associate emotions with the audio features. By combining the coefficients with the PDF emotions can be predicted.

8. Image is a very powerful means for conveying emotion. Assuming the linguistic information from the images are a good lead for predicting emotions, two high level features are focused and they are the object and the background. An object is important as it is useful in defining an image. Through experiment results [8] it is suggested that there is strong correlation between the emotion and the objects in an image. A feed forward deep neural network is built by combining different feature levels.

   This neural network, given an image produces the value for an emotion. The output emotion values in framework are continuous values in 2D space which are proven to be more effective than using a small number of emotion categories to describe emotions.Thoughthe proposed system finds out objects the appearing in the picture it is not able to characterize the state of the object. Experiment results confirm the effectiveness of the framework in predicting the emotions ofimages.

9. This paper [9] proposes a music suggestion system that learns the feeling of an individual from the signs got by means of physiological sensors. Specifically, the feeling of a client is identified by a gadget which is coordinated with a PPG and GSR sensors. These results are then fed to a recommendation engine as additional information. Results are gotten on thirty two subjects' GSR and PPG signals using various deep learning algorithms.

   When tested on a real time data the results show that the above method provides far better accuracy compared to other existing system which can fit into any model regardless of the recommendation engine. 71% accuracy was obtained for both VA predictions. Similarly, 70% accuracy was obtained for both VA predictions when photoplethysmography signal was used. When both photoplethysmography signals and galvanic skin response signals were combined 72% accuracy was obtained for VA predictions.

10. Manual isolation of a playlist and annotation of songs, with respect to the user's emotional state is work escalated and tedious. Various algorithms have been proposed to robotize this procedure. Be that as it may, the current algorithms are moderate, increase the general expense of the framework by utilizing extra equipment and have less exactness. This paper [10] introduces an algorithm that computerizes the way toward creating an audio playlist, in light of the appearances of a user, for recovery of time and work, put resources into playing out the procedure physically. The algorithm proposed aims at reducing the overall cost and computational cost of the proposed design. It also focuses on increasing the accuracy of the proposed system. The facial expression recognition system using the proposed algorithm is evaluated by testing the system against user independent and user dependent dataset.

CHAPTER 3

SYSTEM ANALYSIS

3.1 PROBLEM DEFINITION

Music is often considered to be voice of the soul as it makes people emote their feelings no matter what the situation is. An angry person tries to calm himself by listening to music which might calm his nerves. A sad person listens to motivating song which helps him to come out of the depression phase. Music and emotion coexist.

Our main objective is to

- Accurately detect the mood of the person
- To create a playlist according to the identified emotion by using a real time dataset.

Real time dataset allows us to capture the person's image at the particular instant based on which songs can be suggested which complies with his mood.

3.2 PROPOSED SYSTEM

In this paper, our main focus is on creating a playlist based on the mood of the individual. Using a camera and by getting the permission of the user, various image of the user is captured at that particular instant. These images undergo a rigorous testing and training process to find out the mood of the individual.

Once the images are captured, they go through various stages like

1. Preprocessing
2. Segmentation
3. Feature extraction
4. Emotion classification
5. Web service integration

At each stage these images go through various filtering process to extract the exact mood of the individual. Features like mouth, eyebrows and eyes are extracted based on which various calculations are done to find the emotion of the person.

A deep learning algorithm called Convolution Neural Network (CNN) is applied to classify the various emotions. Each emotion is associated with a value and when the extracted value of the image falls within the range of the defined values of each emotion, person is said to be in that particular emotional state.

By identifying the emotion of an individual, suitable music can be recommended. The system aims at examining the data provided by identifying the emotion of the user. The proposed system has delivered results with significant accuracy and it also paves the way for further research in this area.

3.3 ADVANTAGES OF PROPOSED SYSTEM

1. Existing models use support vector machine (SVM) algorithm for emotion classification whereas we use convolution neural network (CNN) which is more advanced and effective compared to SVM.
2. One of the most unique features of CNN is its ability to recognize the most important features in an image without any help from humans. They are non-linear and were developed specially to recognize patterns, features within an image with high accuracy.
3. The hidden layers used in CNN are proven to be more effective because as we add more layers the model complexity increases which yields better results. Every image is converted to an array of pixels which is not the case with SVM.

SVM was best-known method for image classification before CNN was introduced. SVM is a supervised learning method applied for data classification. They are best known for their regression analysis. They are linear classifiers whose complexity increases as the number of training dataset increases. An accuracy of 46.74% was attained when 70% of the data was used for testing.

3.4 REQUIREMENTS

3.4.1 FUNCTIONAL REQUIREMENTS

In order for every software application to run properly, it needs to satisfy a lot of functions that are to be deployed in it. These functions are nothing but various operations that are performed in each step while developing the application. This step comes under the best practices of developing an application. Functional and Non-Functional Requirements together set a list of rules that govern the smooth running of an application and it also helps the developer and the user to determine the software and hardware requirements that are needed to run the application. Functional Requirements that are required to run this application is

Python: Python programming language was developed in the year 1991 by Guido Van Rossum. The syntaxes used in the language makes it very comfortable and easier for developers to work with. Because of this very reason, this programming language can be used both in small and large scale. They are dynamic and garbage collected.

OpenCV: It is a python-based library specifically used in image processing. It consists of many computer vision algorithms. It's a cross platform library and supports many deep learning algorithms and frameworks such as caffe, PyTorch, TensorFlow etc. Apart from image processing, OpenCV is also used for

- Video analysis
- Camera calibration
- Object detection
- Core functionality
- High-level GUI

Keras: Keras was developed by Francois Chollet which is a high-level neural network API. It can run on top of Theano, CNTK, TensorFlow etc. The Keras API is modular, user-friendly and extensible. The two most used Keras Models are

- Sequential
- Functional

3.4.2 NON-FUNCTIONAL REQUIREMENTS

Non-functional requirements are used to set conditions to monitor the performance characteristic of the application. It describes how a specific function in the application works. They also determine the overall quality of the project and hence it is a very important aspect in any software development process.

The Non-Functional Requirements include

- Usability: It refers to the easiness of the application and determines the ease with which it can be used by the user. Usability can be said to be high when the knowledge required to use the application is less and the efficiency of its functionality is high. It is also a main criterion which can determine the satisfaction of the user.

- Accuracy: Accuracy determines the relative closeness of the value produced by the system to that of the ideal value. Less the difference between the system value and the ideal value, more is the accuracy. It is also one way to determine how the application works better compared to the other similar applications.

- Responsiveness: Responsiveness is determined by completing the software operations with minimal errors or no errors. It is directly proportional to the stability and the performance of the application. The Robustness and Recoverability can also be determined by this criterion.

- Scalability: Scalability is used to determine the growth of the project. It determines how much room the application can have in order to include more features in the future. It determines the sustainability of the project. This is one criterion which is used to develop long-term models for business growth.

3.4.3 HARDWARE REQUIREMENTS

Processor: Intel I3 processor

Storage Space: 500 GB.

Screen size: 15” LED

Devices Required: Web camera, Mouse and a Keyboard

Minimum Ram: 4GB and a good Internet connection.

3.4.4 SOFTWARE REQUIREMENTS

OS: Windows 7 and above /UBUNTU

Programming Language: Python

Software: JetBrains PyCharm Community Edition 2017.1.4 x64

Backend: Keras

Additional requirements: TensorFlow

3.5 ISSUES IN EXISTING METHODOLOGY

- As mentioned above existing methodology use SVM algorithm for data classification. Only 46.74% was attained when 70% of the data was tested. They need high parameter tuning since they are non-parametric models.
- In the existing system, the emotion of the user is not taken into consideration.
- Recent issues involve multiple frequency estimation and music similarity computation.
- A QBSH (Query by singing and humming) system determines a song by using its contents (tune and rhythm). But the issue revolving around this option is that they are time-consuming and not always satisfies the user.

3.6 NEW METHODOLOGY

3.6.1 GATHERING DATASETS

Real time datasets are used for emotion classification by using a web camera or mobile camera. Images are captured after getting permission from the user. The captured images are compared with FER 2013 datasets for emotion classification. FER 2013 datasets consist 37887 grey scale images with 7 different emotions where 0 is Anger, 1 is Disgust, 2 is Fear, 3 is Happy, 4 is Sad, 5 is Surprise and 6 is Neutral.

3.6.2 FACIAL DETECTION AND RECOGNITION

Facial detection is the process of identifying a human face within an image. If there are five faces in an image face detector must be able to detect all the five faces. Haar's cascade is used for this purpose. It works often by searching for human eye i.e. a valley region.

Haar Cascade

- This method gives the output as one if the region show any feature or object and zero when no object is detected. By moving the search window an individual can search every location for identifying objects and features.

- The classifier is structured with the goal that it tends to be effectively resized so as to have the capacity to discover the objects of various sizes, which is more proficient than resizing the picture itself. Thus, to discover an object of an obscure size in the picture the output strategy ought to be completed a few times at various scale**s**.

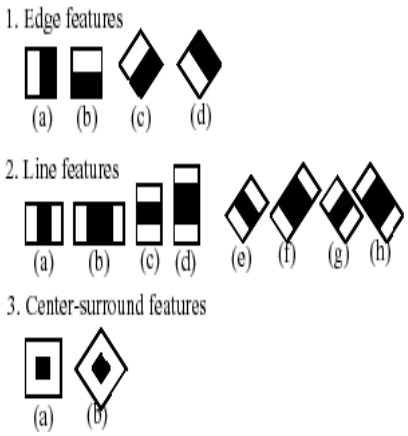

Fig 3.1 Haar-Cascade Classifier

In order to find one person from the five people facial recognition is used. It does something beyond just recognizing faces. It uses a biometric technology that snaps the image of the human's face and compares it with the existing images that are already stored in a database. It will be able to identify the person to whom the face belongs to. LBPH face recognizer is used for this purpose. LBPH face recognizer extracts the face, crops them and resizes them. Finally, these processed images are converted to grey scale.

CHAPTER 4

SYSTEM DESIGN

4.1 ALGORITHM

4.1.1 CNN Algorithm

Convolution Neutral Network as it is called was developed mainly to recognize patterns, features within an image with high accuracy. Another advantage of CNN is its ability to recognize the most important features in an image without any help from humans. A CNN consists of two parts: feature extraction part and the classification part. Feature extraction is performed by convolution and pooling layers.

4.2 MODULES

1. Image Capturing
2. Pre-Processing
3. Segmentation
4. Feature extraction
5. Emotion Classification

Fig 4.1 Basic and Combined Emotions

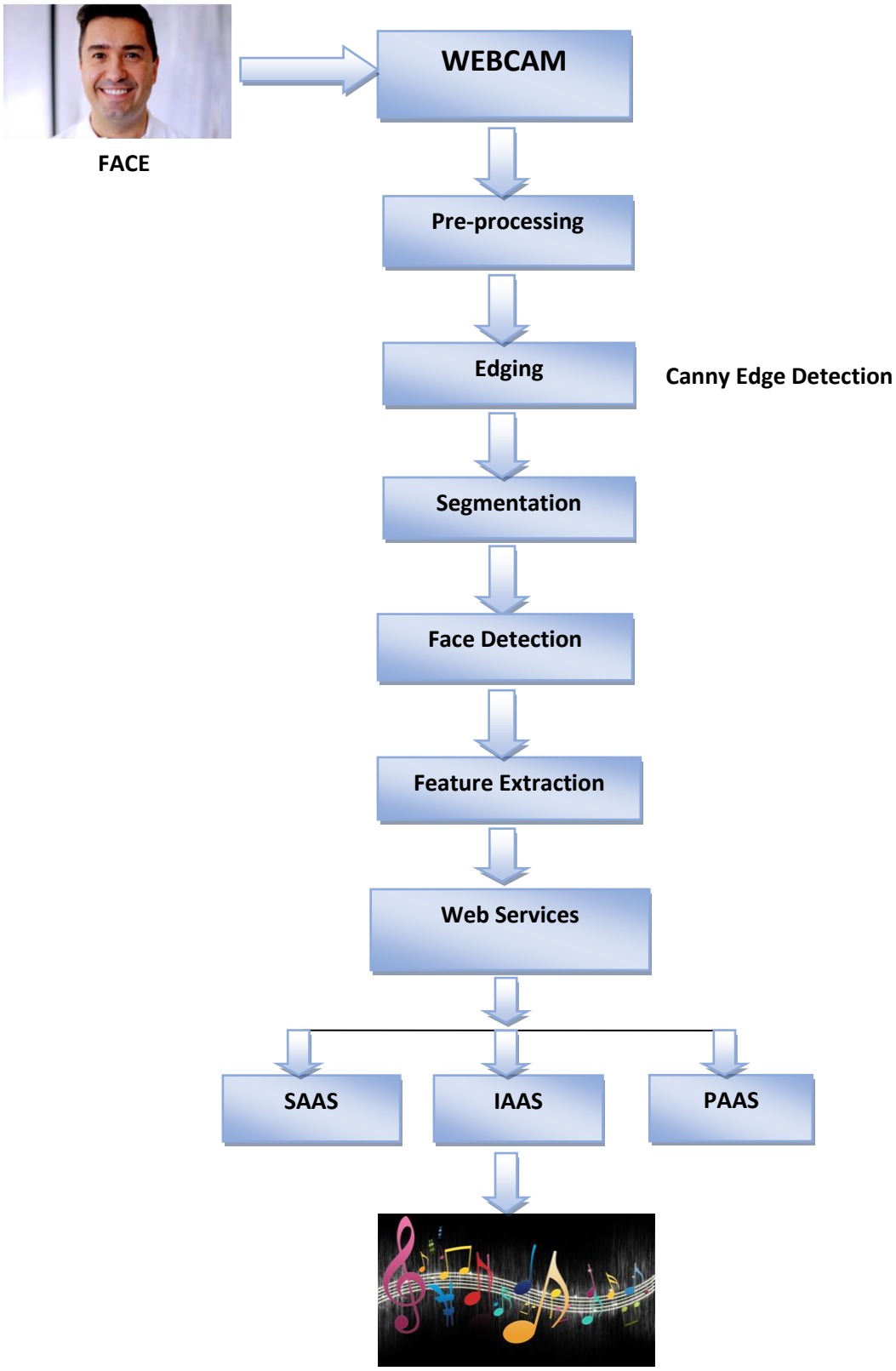

Fig 4.2 Block diagram of the application

4.3 FUNCTIONALITIES

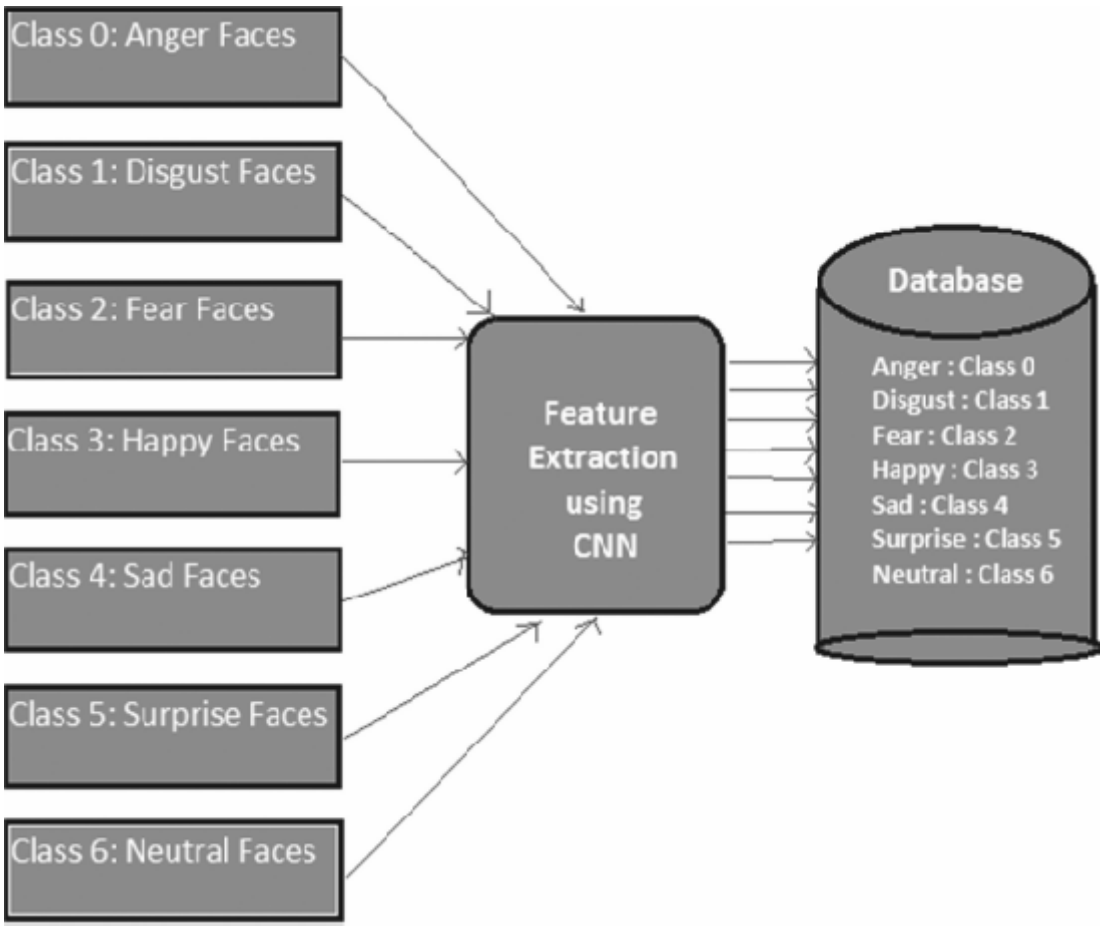

Fig 3.2 Proposed CNN Architecture

4.4 UNIFIED MODELING LANGUAGE

UML is a modelling language which is used to visualize a way the software has been developed. It is a visual language which consists of serious of steps in which the application works. The UML diagrams depict the structure and behavior of the software application.

Visualizing a project before developing an application makes the job easier. It can be used to make a list of requirements for software development. Also, a lot of time is saved where there is a definite model to look up to while developing a software.

UML Diagrams can be broadly classified into:

- Behavior diagrams: They capture or visualize the behavior of the diagram. They depict how the behavior changes w.r.t each task or function. These diagrams include Interaction diagrams, Use-case diagrams, State diagrams, Activity Diagrams etc.
- Structural diagrams: They concentrate on the structure of the application rather than the functionalities. These diagrams include Class Diagrams, Deployment diagrams, Component diagrams etc.

4.4.1 USE CASE DIAGRAM

They are generally used to visualize the various tasks that are performed in the application. They also depict the different users who have access to perform these tasks. Use-case diagrams come under behavior diagrams because of its emphasis on the tasks performed and the users(actors) who perform these tasks. The various tasks that are performed in the application is

1. Getting permission for webcam
2. Image pre-processing
3. Image segmentation
4. Feature extraction
5. Emotion classification
6. Music recommendation based on the mood

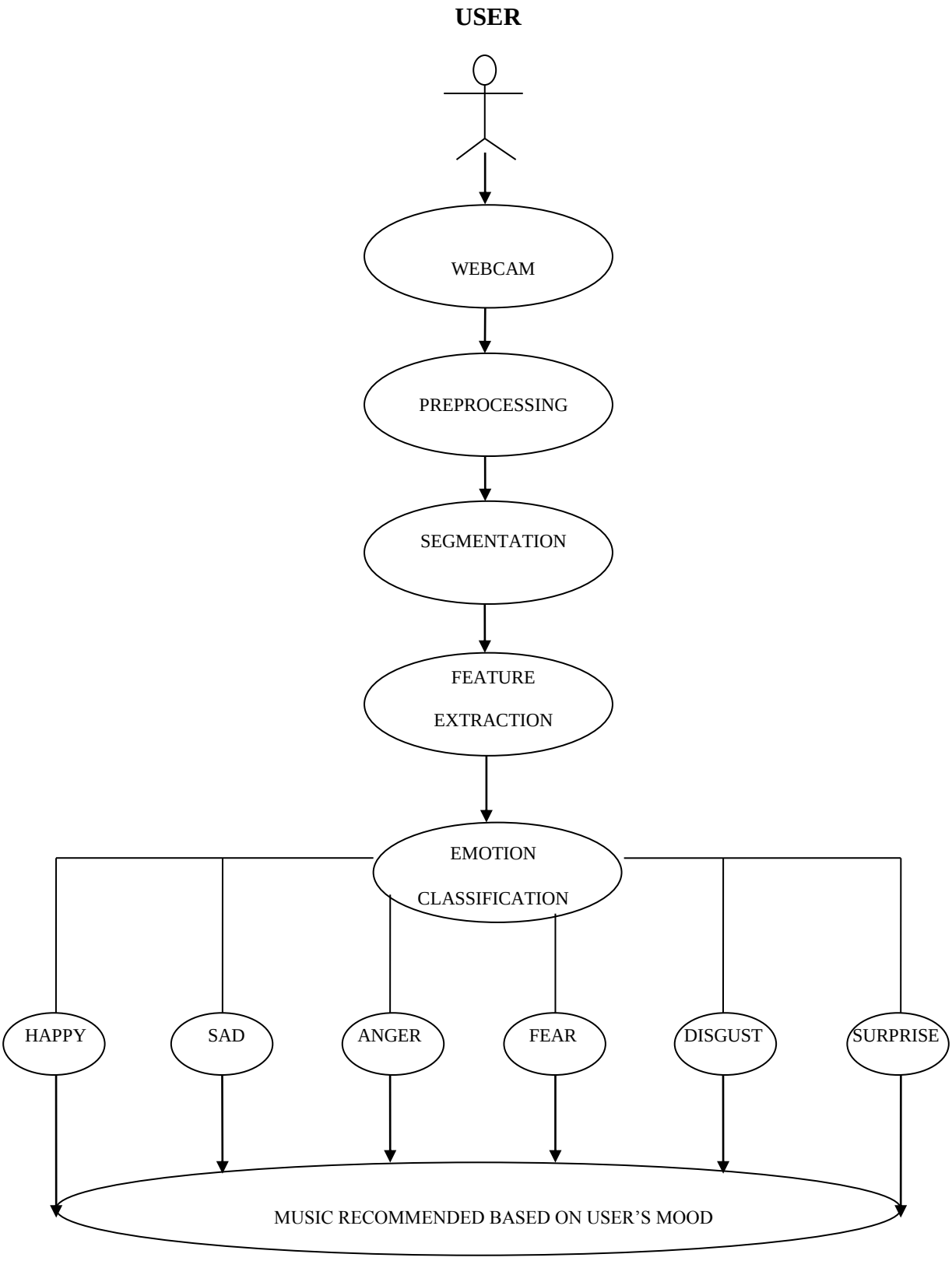

Fig 4.3 The above diagram depicts the various tasks performed in the application and the actor (the user) who uses this application.

4.4.2 ACTIVITY DIAGRAM

Activity diagram depicts the flow or the sequence of control in an application. We can actually use the activity diagram to verify every task that is performed in the use-case diagram. It also depicts the steps of execution. They basically depict the workflows in the software application. It also emphasizes on the sequence of tasks and the conditions that are to be met in order for a particular task to perform. This way, it gives us information about what causes a particular task to happen. So, this gives us a high-level visualization of the application. The main objectives are

- Depict the activity flow
- Describe the sequence (branched or sequential)

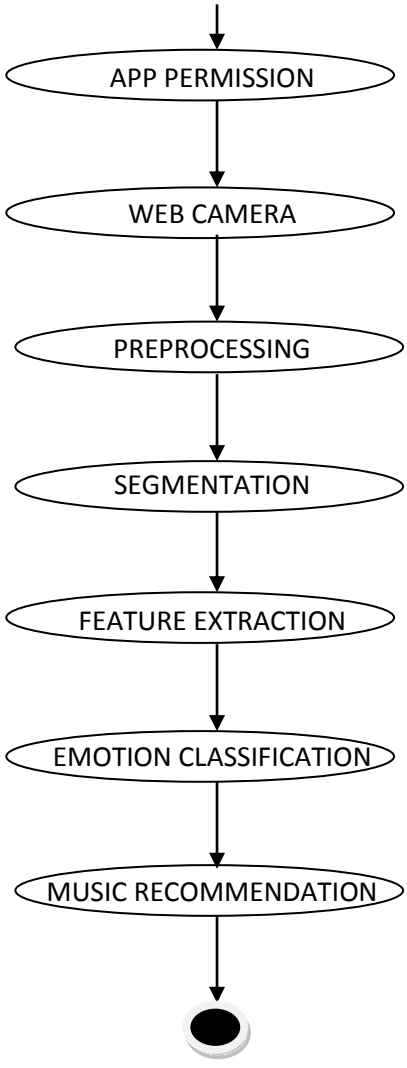

Fig 4.4 Activity Diagram

4.4.3 SEQUENCE DIAGRAM

It is used to depict the interaction among the objects in an application. They also describe the changes in the behaviour of these objects after they interact with each other. They also describe the order in which these interactions take place. Also, developers use these diagrams to make note of all the requirements that are to be needed in order to develop the application.

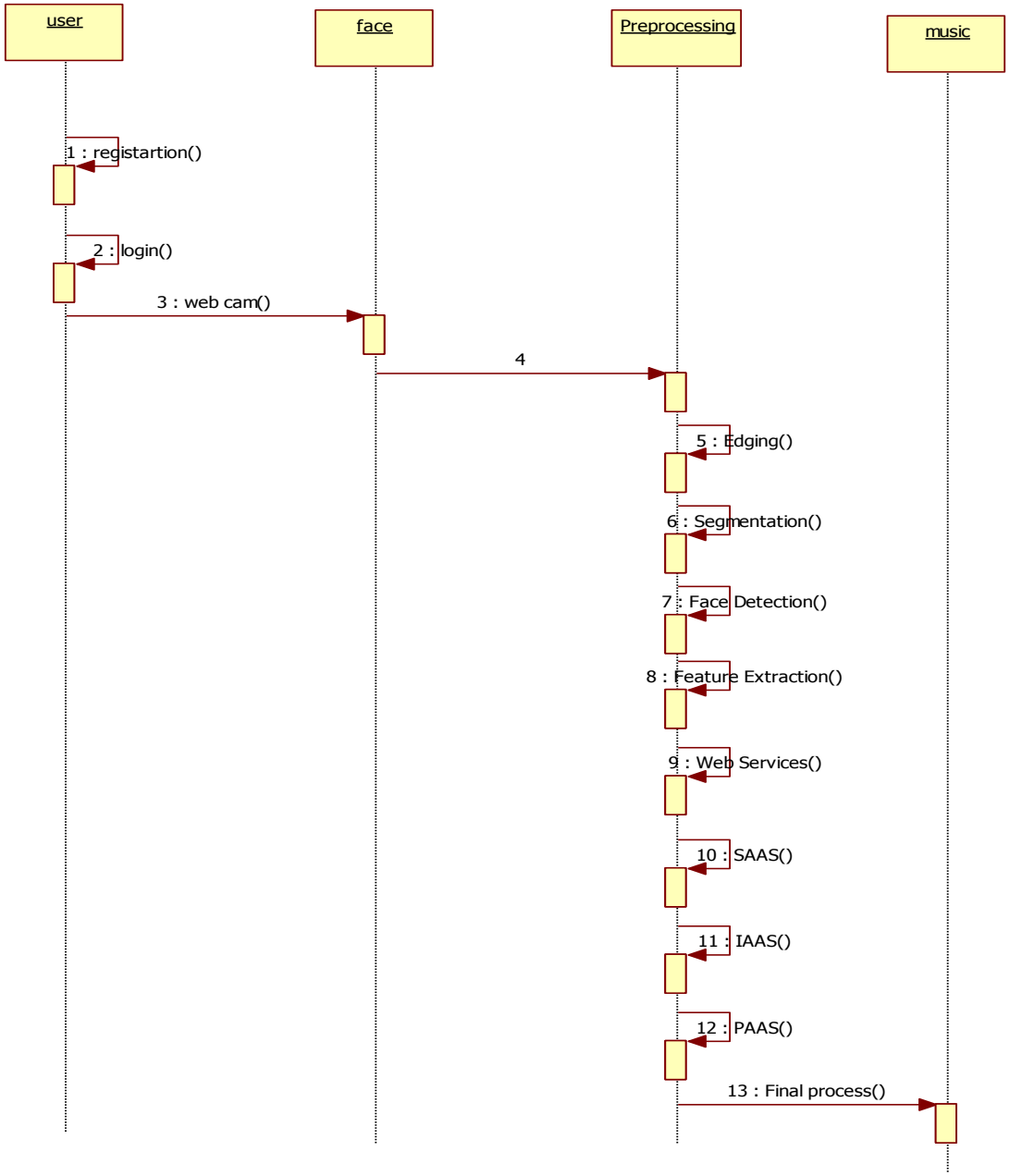

Fig 4.5 Sequence diagram

4.4.4 COLLABORATION DIAGRAM

Collaboration diagram is also known as the communication diagram. As the name suggests, it depicts or describes the sequence messages that are shared among the objects in a software application. They also primarily focus on the objects and the interactions among them. Objects and links are represented in free-form in this diagram.

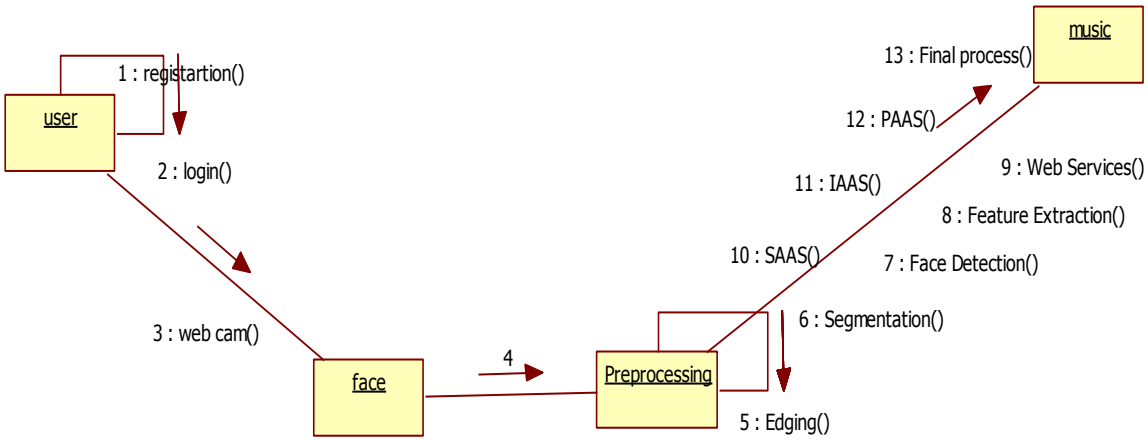

Fig 4.6 Communication diagram

CHAPTER 5

IMPLEMENTATION

5.1 CASCADE CLASSIFIER

The process of detecting a face within an image is facial detection. If an image face detector has five faces, it must be capable of detecting all five faces. For this purpose, the hair cascade is used. It often works by looking for a human eye, i.e. a region of the valley.

Haar Cascade

The classifier is structured with the goal that it tends to be effectively resized so as to have the capacity to discover the objects of various sizes, which is more proficient than resizing the picture itself. Thus, to discover an object of an obscure size in the picture the output strategy ought to be completed a few times at various scales.

If the region is likely to display the features, the classifier will give the output as zero or otherwise if there is any discrepancy, it will give the output as one. After converting it into the pixel format, the classifier will go across the image i.e., go by pixel by pixel to consolidate the entire image and will take note of its location.

LBHP Face Recognizer

The facial recognition is used to find one person from the five people. It does something beyond the mere recognition of faces. It uses a bio-metric technology that captures and compares the digital image of the face of an individual to images stored in a database. The person to whom the face belongs will be able to be identified.

For this purpose, the LBPH face recognizer is used. The face recognizer of LBPH extracts the face, crops it and resizes it. These processed images are finally converted to a grey scale.

Therefore, we can conclude by saying that all face recognizers are face detectors but not all face detectors are face recognizers.

5.2 DATASETS

Only real-time datasets are used for classification. The images are gathered by asking app permissions through webcam or mobile cam depending upon where the application is being ran. Images are gathered after getting these permissions only which does not violate any privacy policy.

We make use of Kaggle's FER 2013 datasets for comparing the obtained images for emotion classification. The FER 2013 dataset consists of over 38,000 grey scale images with seven different emotions. Every emotion is coded such as 0 for Anger, 1 for Disgust, 2 for Fear, 3 for Happy, 4 for Sad, 5 for Surprise and 6 for Neutral.

5.3 IMAGE PRE-PROCESSING

In the overall process, pre-processing plays a key role. Pre-processing stage improves image quality and locates interesting data by removing noise and smoothing the image. It removes redundancy without the details of the image. Pre-processing also includes image filtering and normalization, which results in uniform image size and rotation.

Normalization: Let's assume that the image of the person is somewhat whitened up because of the light from the background. In that case, the face will not be clear and classification will become very difficult and the accuracy will be affected very badly. So, in this case we need to somehow remove the whites in the background and enhance the face of the person. This is called as Normalization.

Edging: After we perform normalization, we have to identify the shape of the object that is needed. We need to eliminate all the unwanted edges in the picture. This is performed by traversing through the horizontal and vertical axes to find the maximum pixel. The boundary which is created by the maximum pixels in the vertical and horizontal axes are said to be the edges and all the others edges are neglected.

Data Augmentation: Sometimes in order to achieve better accuracy, we collect all the data possible. For e.g., suppose if the person has tilted his face more towards the right, we cannot get the values from the left side of his face. And, this is where data augmentation comes into

play. Since the face of any human is perfectly symmetrical, we use that logic to get the values from the left side of his face.

Also, we perform automatic tilting, cropping and zooming of images to get the best possible values. All this will help us to achieve the best ever accuracy possible.

5.4 SEGMENTATION

Based on parameters like texture, intensity and edges, the images are segregated and this process is called as Image Segmentation. This is a very important step to be performed in Image classification. The image is segregated into homogeneous regions with respect to the above three parameters. This process adds to the motive of getting better accuracy during the emotion classification.

5.5 FEATURE EXTRACTION

To find the emotion of a person, it is important to identify those features which brings a difference to the face while showing different emotions. These features are eyes, eyebrows and mouth. With the change in movement of these features, we can identify the emotion of the person.

For e.g., if the person has his mouth opened and the size of his eyes larger than usual, then we can consider him to be surprised.

The following facial features are extracted:

Eye: Human eyes show solid vertical edges because of its iris and white eye. In this manner, the Sobel cover is connected to a picture and the even projection of vertical edges can be acquired to decide the vertical facilitate of the eyes.

Eyebrow: 2 rectangular areas in the edge picture which lies over the eye areas are chosen as the eyebrow locales. The edge pictures of these two regions are acquired for further enhancement. Presently sobel technique was utilized in acquiring the edge picture since it has the ability to distinguish a larger number of edges than Roberts strategy. These pictures are

then enlarged to fill the gaps. The resulting pictures are utilized in refining the eyebrow locales.

Mouth: To calculate centroid of the mouth the left, right, top and bottom most of the points are extracted.

5.6 EMOTION CLASSIFICATION

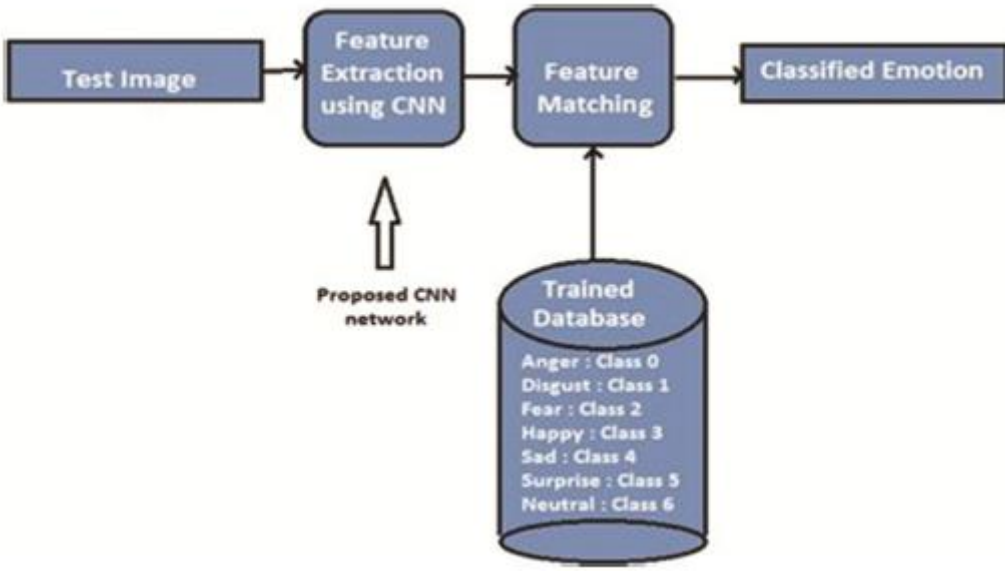

Fig 5.1 Proposed testing architecture

After the image has undergone pre-processing, we can consider the image to be ready for classification. The image is previously converted into the pixel format before we send it to the input layer of the CNN architecture.

INPUT LAYER

- The pre-processed image is sent as pixel format into this layer.
- It consists of three constraints i.e., (w x h x c) where w stands for width, h stands for height and c stands for the colour channel used. In our case, it is (48 x 48 x 1), the colour channel is 1 which denotes grey-scaled images.

CONVOLUTIONAL LAYER

- This layer is used to develop a feature map which consists of various feature identifiers which are used to enhance the image for better classification.
- A dot product is calculated between the input layer neurons and weights that are assigned to each filter. This filter is also called as the kernel.
- Here the hyper-parameter would be the number of filters that the image passes through.

MAX POOLING LAYER

- The next layer after the convolutional layer is the Max-Pooling layer.
- This layer is used to reduce the computing time and increase the efficiency.
- The size of the feature map developed in convolutional layer is reduced in order to retain the maximum pixel value.

DENSE LAYER

- It is also called as fully connected layer. It can also be considered as a part of the output layer.
- They are mostly utilized during the last phase of CNN model
- They recognize all the vital features that actually draw out the entire picture.

OUTPUT LAYER

- This is the last layer of a CNN model and is connected to the dense layer.
- Once the probabilities are calculated they are displayed as output.

CHAPTER 6

SOFTWARE TESTING

6.1 UNIT TESTING

It is the process of testing each and every module developed by the developers. The entire program is segmented into many packages which consist of small units of code. It improves the overall design of the module and refactors the code wherever necessary. These modules are tested independently irrespective of other modules. They are tested in a sequential order and it checks for redundancy. In case of redundancy it deletes the duplicate records. It also checks for run time error and checks if the link provided take them to the respective page. Advantage of performing unit testing is its ability to check each module individually which is helpful in finding the smallest of smallest errors. Since unit testing is done at a very early phase the cost of testing is minimal when compared to other testing. Modules which are too big for unit testing can be evaluated using integration testing.

6.2 INTEGRATION TESTING

This is next step after unit testing is performed. Once, each module tested independently is clear of errors, these individual modules are combined together and tested as a whole. The main reason for performing this test is to check for faults when all the units are combined. There are different ways in which these units can be integrated. They are

1. Top Down Integration -Top-down integration combines and tests all the modules from top to bottom. But one disadvantage of this testing is that it needs more stubs.
2. Bottom Up Integration -The bottom up approach is the vice-versa of top-down approach. Important modules are tested last which can create problems during integration.

3. Big-Bang Integration - In this form of testing all the functionalities are integrated and tested simultaneously. This approach is dependent on the number of modules present. Lesser The modules more effective it is.
4. Hybrid Integration – It is a combination of all the above approaches.

6.3 SYSTEM TESTING

System Testing is the next step after integration testing. In this process the whole product is tested for faults and errors. They are of two types:

1. Black box testing
2. White box testing

An example for this is manufacturing of ballpoint pen. The cap, the ink cartridge, the body, the tail is produced separately and tested separately (unit testing). When two or more modules are ready, they are combined and Integration Testing is done. When the complete pen is assembled, System Testing is done. It considers the whole system as single entity.

1. Black Box Testing

It is a testing procedure which is carried out by the testers. This software can be tested without knowing the internal structure of program. Programming Knowledge is not needed to carry out this form of testing procedure.Its main ambition is to check for the operation that is performed by the system. It is less time consuming. Black box testing is otherwise called functional test or external testing. It is not preferable for algorithm testing. It can be tested on supreme levels of testing like acceptance testing.

2. White Box Testing

It is a testing procedure which is carried out by s/w developers. The functionality of the program must be known to the developer. Programming knowledge is a must to perform White Box Testing. It is otherwise called interior testing or structural testing. Its main aim is to check program code, loops, conditions, branches and how system is performing. It can be tested on higher levels of testing like acceptance testing and acceptance testing.

6.4 REGRESSION TESTING

This is one of the most important type of testing when it comes to the proper development of a software. We can also consider it as one important step in the Software Development Life Cycle (SDLC). Every software has a certain type of functionalities which needs to be updated every time. This is usually done to ensure its stability in all platforms. So, for this to be ensured, these functionalities have to updated with new piece of code every time. Therefore, in order to ensure that the new code doesn't affect the new functionality, regression testing is carried out. This is usually done by experts or software developers who have deep understanding of the software operations in and out.

6.5 SMOKE TESTING

It is also one aspect to make sure that the functionality is just working fine irrespective of the new code that is added to alter it. One of the most important reason to perform this form of testing is to remove all those lines of code that is not required anymore and see to it that they do not affect the functionality of the software. It covers most of the crucial functions of the software but does not analyse them in detail. The consequence of this test is utilized to choose whether to continue with further testing. In the event that the smoke test passes, proceed with further testing. On the off chance that it comes up short, end further tests and request another form with the required fixes.

6.6 ACCEPTANCE TESTING

This is the final phase of testing which is performed by or in front of clients. This testing is mainly done to check whether the developed product satisfies the client's requirement.

They are 4 different ways in which acceptance testing can be performed. They are:

1. User acceptance testing
2. Business acceptance testing
3. Alpha testing
4. Beta testing

CHAPTER 7

RESULTS

7.1 OBTAINED RESULTS

After the image classification, we could successfully and accurately identify the emotion of the person.

The different emotions are Angry, Sad, Surprised and Happiness.

Output

Fig 7.1 Angry

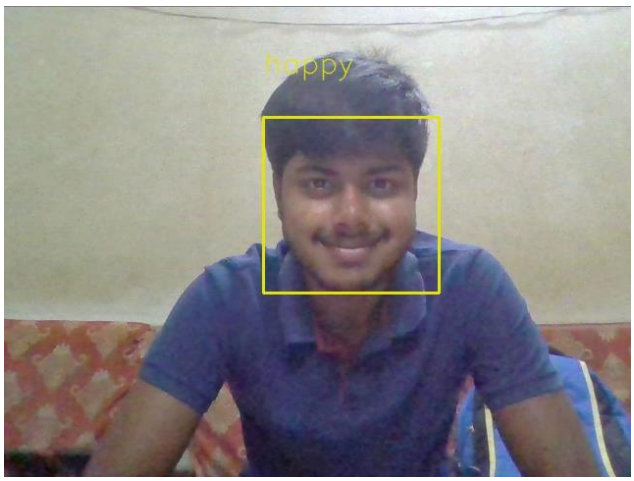

Fig 7.2 Happy

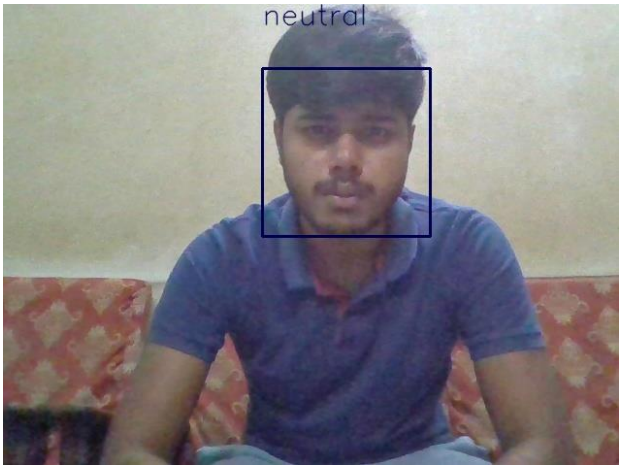

Fig 7.3 Neutral

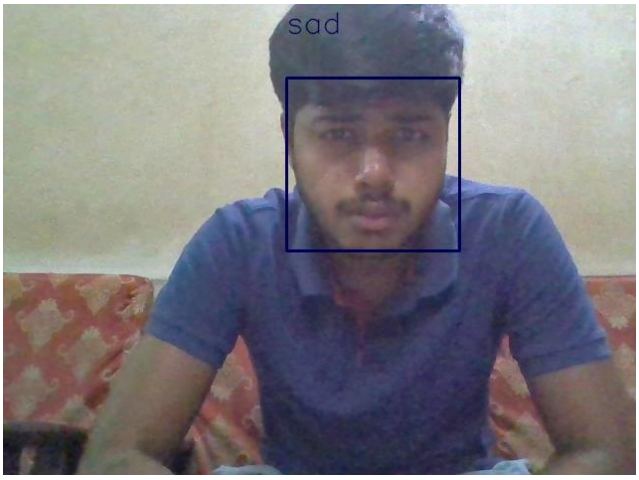

Fig 7.4 Sad

7.2 NORMALIZED CONFUSION MATRIX

We can summarize the performance of a classification model using a technique called as the confusion matrix. By summarizing, we mean that it helps us to understand where the classification model is getting it right and where it is getting it wrong and what kind of error it makes. Basically, we can call it as the summary of all the prediction or classification results which we obtain on a problem. It takes the count of all the correct and incorrect predictions and this acts as the key to the confusion matrix.

The confusion matrix shows how confused your model of classification is when predictions are made. It not only tells about the errors made by the classifier, but also the type of errors it generates. Below is the normalized confusion matrix of the classifier we have used

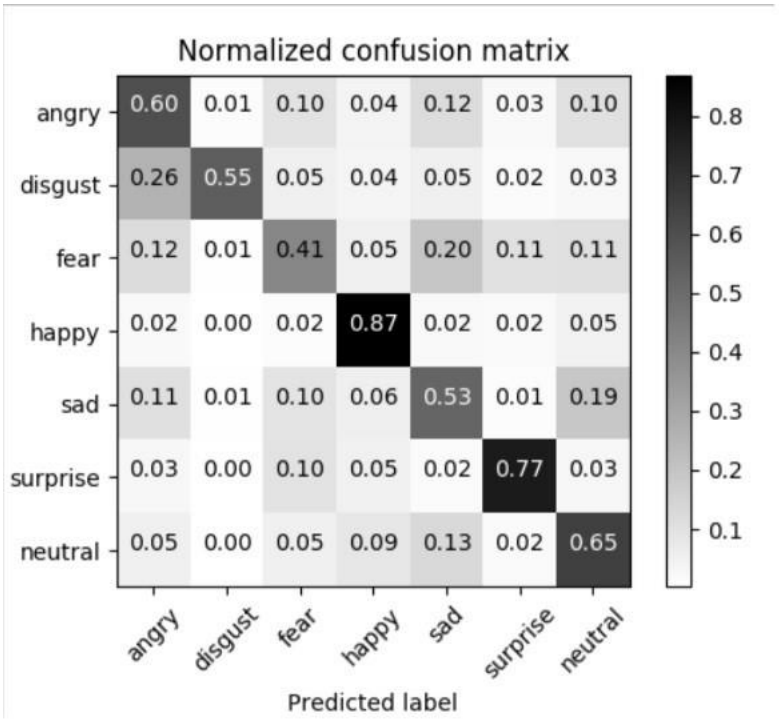

Fig 7.5 Normalized Confusion Matrix

In the above matrix, the diagonal elements contain values which are true and correct that has been predicted by the classifier. The non-diagonal elements represent all the negative values i.e., the values which have errors while the classifier is performing the classification process. The row values represent the instance of the actual class whereas the column represents the instance of the predicted class.

CONCLUSION

Since music has the power to emote user's feelings a generic model is implemented to recommend music based on the user emotions. Human emotions play an important in expressing the thought of an individual. The main goal of the system is to detect changes in the emotional state of the user and play music according to the user's preferences by exploring various music tracks. The system uses CNN algorithm for emotion classification which can be determined by change in shape, size and movement of eyebrows, eyes and mouth. They fall into one of the six basic types of emotions which are sadness, happiness, anger, fear, disgust and surprise based on which a playlist is generated. The main reason for using CNN algorithm over SVM is its ability to recognize the most important features in an image without any help from humans. Also, SVM's prediction accuracy is found to be less when compared to CNN's accuracy. The proposed system has delivered results with significant accuracy. Since human emotions are not consistent and they are actually a result of internal and external circumstances happening around an individual it is difficult to get 100% accuracy. But with better algorithm and intense research a perfect emotion-based music recommendation system can be developed.The Proposed system is tested against a web camera. The total cost involved in implementing this project is almost negligible. Average estimated time for various modules of proposed system.

i. Every one of those tunes that are energetic, playful and cheerful are classified under joy.

ii. Sad and depressing songs are classified under the class sad. Songs that reflect mentality, 'anger related with patriotism', and are vengeful are grouped under anger

iii. The Joy-Anger classification is related with tunes that have outrage in an energetic mode.Sad-anger category comprises of all those songs that revolve around the theme of being extremely depressed and angry.

iv. When a user is in fear or surprised, songs from the other category are suggested.

FUTURE SCOPE

Emotion Classification has been the area of research which promises to have a wide range of applications. Before performing emotion classification, the facial recognition step is very important. Recently Apple has brought in the technology of facial recognition for unlocking apple devices. But it has the scope for performing facial authentication in much higher levels. We can consider Facial Recognition to be one of the most effective biometric sensors which shows a lot of potential. By enforcing it in higher levels, we can cite the example of using them in ATMs or ensuring authorization in offices. It leads to much better resource integrity and safety.

When it comes to emotion recognition, marketing researchers are looking for ways to add values to their brands. Their research involves collecting data from their clients, studying their emotions which in turn can help them to deliver goods to their customers with a personal touch. This method enhances customer loyalty. In the future, systems which uses emotion recognition can answer automatically instead of asking questions to their customers. Emotion recognition will be seen a valid business tool with great potential.

When we enter into any major supermarkets, we are asked to press a button to show our feedback. Instead, supermarkets can use this technology for scanning their faces and detecting their emotions while they are standing in queues. This data can help the supermarket for further providing services to their customers.

One of the main reasons for accidents these days is drunk and drive cases. Alcohol leads to severe instability to the driver which causes him to lose focus while driving. Our technology can be used as a warning system to alert the driver that it is not safe to drive. Or, it can also be used as a technology in automated cars which can help the driver in reaching his destination without actually having him to drive it.

Thus, both facial recognition and emotion detection has shown a lot of potential and it is us who have to tap this potential and make best use of it.

REFERENCES

1. Anagha S. Dhavalikar and Dr. R. K. Kulkarni “Face Detection and Facial Expression Recognition System “Institute of Electrical and Electronics Engineers (IEEE 2014)

2. Byeong-jun Han &Seungmin Rho &Sanghoon Jun &EenjunHwang”Music emotion classification and context-basedmusicrecommendation” Springer Science + Business Media, LLC 2009

3. Renata Lopes Rosa, DemóstenesZegarra Rodríguez and GraçaBressan“Music Recommendation System Based on User’s Sentiments Extracted from Social Networks”2015 IEEE International Conference on Consumer Electronics (ICCE)

4. Yajie Hu and MitsunoriOgihara “NEXTONE PLAYER: A MUSIC RECOMMENDATION SYSTEM BASEDON USER BEHAVIOUR”12th International Society for Music Information Retrieval Conference (ISMIR 2011)

5. Kyoungro Yoon, Senior Member, IEEE, Jonghyung Lee, and Min-Uk Kim “Music Recommendation System Using Emotion Triggering Low-level Features”IEEE Transactions on Consumer Electronics 2012

6. Fang-Fei Kuo1, Meng-Fen Chiang2, Man-Kwan Shan2 and Suh-Yin Lee “Emotion-based Music Recommendation by Association Discovery from FilmMusic” 2005

7. Yu-Hao Chin, Jia-Ching Wang, Senior Member, IEEE, Ju-Chiang Wang, and Yi-Hsuan Yang, Member, IEEE”Predicting the Probability Density Function of Music Emotion using EmotionSpace Mapping” IEEE 2018

8. Hye-Rin Kim, Yeong-Seok Kim, SeonJoo Kim, In-Kwon Le”Building Emotional Machines: Recognizing Image Emotions through DeepNeural Networks”IEEE TRANSACTIONS ON MULTIMEDIA 2018

9. DegerAyata, Yusuf Yaslan and Mustafa E. Kamasak "Emotion Based Music Recommendation System Using Wearable PhysiologicalSensors"IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, 2018

10. AnukritiDureha "An Accurate Algorithm for Generating a Music Playlist based on Facial Expressions "IJCA 2014

11. AnaghaS.Dhavalikar and Dr. R. K. Kulkarni, "Face Detection and Facial Expression Recognition System" 2014 International Conference on Electronics and Communication System (ICECS -2014).

12. Yong-Hwan Lee , Woori Han and Youngseop Kim, "Emotional Recognition from Facial Expression Analysis using Bezier Curve Fitting" 2013 16th International Conference on Network-Based Information Systems.

13. ArtoLehtiniemi and Jukka Holm, "Using Animated Mood Pictures in Music Recommendation", 2012 16th International Conference on Information Visualisation.

14. F. Abdat, C. Maaoui and A. Pruski, "Human-computer interaction using emotion recognition from facial expression", 2011 UKSim 5th European Symposium on Computer Modelling and Simulation.

15. T.-H. Wang and J.-J.J. Lien, "Facial Expression Recognition System Based on Rigid and Non-Rigid Motion Separation and 3D Pose Estimation," J. Pattern Recognition, vol. 42, no. 5, pp. 962-977, 2009.

16. Renuka R. Londhe, Dr.Vrushshen P. Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach", International Journal of Soft Computing and Engineering (IJSCE) Volume-2, May 2012.

17. S. Yang and B. Bhanu, "Facial Expression Recognition Using Emotion Avatar Image," 2011, pp. 866–871.

18. M. Bejani, D. Gharavian, and N. M. Charkari, "Audiovisual emotion recognition using ANOVA feature selection method and multiclassifier neural networks," Neural Comput. Appl., vol. 24, no. 2, pp. 399–412, 2014.

19. https://keras.io/models/about-keras-models/
20. https://keras.io/layers/about-keras-layers/
21. https://keras.io/backend/
22. https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network
23. https://www.mathworks.com/discovery/deep-learning.html
24. http://deeplearning.stanford.edu/tutorial/supervised/FeatureExtractionUsingConvolution/
25. https://ieeexplore.ieee.org/document/8474912/
26. https://appliedmachinelearning.blog/2018/11/28/demonstration-of-facial-emotion-recognition-on-real-time-video-using-cnn-python-keras/
27. https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2
28. https://appliedmachinelearning.blog/2018/11/28/demonstration-of-facial-emotion-recognition-on-real-time-video-using-cnn-python-keras/
29. http://softwaretestingfundamentals.com/acceptance-testing/
30. http://softwaretestingfundamentals.com/smoke-testing/
31. https://searchsoftwarequality.techtarget.com/definition/regression-testing
32. https://www.tutorialspoint.com/software_testing_dictionary/white_box_testing.htm
33. https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8
34. https://en.wikipedia.org/wiki/Convolutional_neural_network
35. https://www.softwaretestingclass.com/difference-between-black-box-testing-and-white-box-testing/
36. https://www.guru99.com/system-testing.html
37. http://www.cs.cmu.edu/~bhiksha/courses/deeplearning/Fall.2016/pdfs/Simard.pdf
38. https://ieeexplore.ieee.org/abstract/document/6165309/
39. http://cis.csuohio.edu/~sschung/CIS660/DeepFaceRecognition_parkhi15.pdf
40. http://openaccess.thecvf.com/content_cvpr_2015/html/Li_A_Convolutional_Neural_2015_CVPR_paper.html

APPENDIX

```
import cv2
import numpyasnump
import winsound


from keras.modelsimportload_model
from statistics import mode
from utils.datasetsimportget_labels
from utils.inferenceimportdetect_faces
from utils.inferenceimportdraw_text
from utils.inferenceimportdraw_bounding_box
from utils.inferenceimportapply_offsets
from utils.inferenceimportload_detection_model
from utils.preprocessorimportpreprocess_input
from tkinterimport*
import tkinterastkin


from tkinterimportfontastfont


from PIL import Image, ImageTk
from pygameimportmixerimporttime
import tkinter


USE_WEBCAM = True
emo_model_path = './models/emotion_model.hdf5'
emo_labels = get_labels('fer2018')

frame_wind = 10
```

```
emo_offsets = (20, 40)
face_cascade = cv2.CascadeClassifier('./models/haarcascade_frontalface_default.xml')
emo_classifier = load_model(emo_model_path)

emo_target_size = emo_classifier.input_shape[1:3]
emo_window = []
cv2.namedWindow('window_frame')
vid_capture = cv2.VideoCapture(0)

cp = None
if (USE_WEBCAM == True):
   cp = cv2.VideoCapture(0)  # Webcam source
else:
   cp = cv2.VideoCapture('./demo/dinner.mp4')  # Video file source

while cp.isOpened():  # True:
ret, bgr_img = cp.read()

gray_img = cv2.cvtColor(bgr_image, cv2.COLOR_BGR2GRAY)
print(gray_image)
rgb_img = cv2.cvtColor(bgr_image, cv2.COLOR_BGR2RGB)
face = face_cascade.detectMultiScale(gray_image, scaleFactor=1.1, minNeighbors=5,
minSize=(30, 30), flags=cv2.CASCADE_SCALE_IMAGE)

for face_coordinatesinfaces:

      x1, x2, y1, y2 = apply_offsets(face_coordinates, emot_offsets)
grey_face = gray_img[y1:y2, x1:x2]
try:
grey_face = cv2.resize(grey_face, (emo_target_size))
except:
continue
```

```
grey_face = preprocess_input(grey_face, True)
grey_face = nump.expand_dims(grey_face, 0)
grey_face = nump.expand_dims(grey_face, -1)
emo_prediction = emo_classifier.predict(grey_face)
emo_probability = nump.max(emo_prediction)
emo_label_arg = nump.argmax(emo_prediction)
emo_text = emo_labels[emo_label_arg]
emo_window.append(emo_text)

if len(emo_window) >frame_window:
emo_window.pop(0)
try:
emo_mode = mode(emo_window)
except:
continue


if emo_text == 'angry' :
print(emo_probability)

if emo_probability>= 0.48181498 and emo_probability<= 0.49352397:

cp.release()
    cv2.destroyAllWindows()
mixer.init()
    mixer.music.load('C:\\Users\\Hariharan\\PycharmProjects\\Emotion-Detction\\Banjaara -
Mohammad Irfan_1417348739963.mp3')
mixer.music.play()

    top = tkinter.Tkin()
canvas = Canvas(top, width=650, height=550)
canvas.pack(fill=BOTH, expand=True)
canvas.create_text(350, 55, text='EMOTION PREDICTON', font=('Helvetica', 21, 'bold'),
justify='center',
```

```
fill='green')
canvas.update
        load = Image.open("./matrix.PNG")
        load = load.resize((378, 378))
        render = ImageTk.PhotoImage(load)

        image = tkin.Label(top, image=render)
image.image = render
image.place(x=20, y=100)

        load = Image.open("./ANG_SMY.png")
        load = load.resize((128, 128))
        render = ImageTk.PhotoImage(load)

        image = tkin.Label(top, image=render)
image.img = render
image.place(x=500, y=250)
top.after(9000, lambda: top.destroy())seconds
top.mainloop()
top.mainloop()



color = emo_probability * nump.asarray((255, 0, 0))
else:
        colour = emo_probability * nump.asarray((0, 255, 0))
colour = colour.astype(int)
    colour = colour.tolist()
draw_bounding_box(face_coordinates, rgb_img, color)
draw_text(face_coordinates, rgb_img, emotion_mode,
color, 0, -45, 1, 1)

bgr_img = cv2.cvtColor(rgb_img, cv2.COLOR_RGB2BGR)
```

```
cv2.imshow('window_frame', bgr_image)
if cv2.waitKey(1) &0xFF == ord('q'):
break

class MusicApp(tkin.Tk):
def __init__(self, *args, **kwargs):
        tkin.Tk.__init__(self, *args, **kwargs)

self.title_font = tfont.Font(family='Helvetica', size=14, weight="bold", slant="italic")
contains = tkin.Frame(self, bg="green")
contains.pack(side="top", fill="both", expand=True)
contains.grid_rowconfigure(0, weight=1)
contains.grid_columnconfigure(0, weight=1)
self.frames = {}
for F in (StartPage, PageOne, PageTwo, PageTree, PageFour, PageFive):
page_name = F.__name__
frame = F(parent=container, controller=self)
self.frames[page_name] = frame

frame.grid(row=0, column=0, sticky="nsew")

self.show_frame("StartPage")
self.centerWindow()
def disp_frame(self, page_name):
'''Show a frame for the given page name'''
frame = self.frames[page_name]
frame.tkraise()
def centreWindow(self):
wid = 650
hei = 550
scw = self.winfo_screenwidth()
sch = self.winfo_screenheight()
            x = (sw - w) / 2
```

```
y = (sh - h) / 2
self.geometry('%dx%d+%d+%d' % (wid, hei, x, y))


class StartPage(tkin.Frame):
def __init__(self, parent, controller):
        tkin.Frame.__init__(self, parent)
self.controller = controller
        label = tkin.Label(self,
text="Emotion",
font=controller.title_font)
label.pack(side="top", fill="x", pady=10)
        label = tkin.Label(self, text="This software recognizes human faces and \ntheir
corresponding emotions from \na video or webcam feed. \nPowered by OpenCV and Deep
Learning. ", font=controller.title_font)
label.pack(side="top", fill="x", pady=10)

btn2 = tkin.Button(self, text="EMO_ON",
command=lambda: controller.show_frame("PageTwo"))
bt5 = tkin.Button(self, text="AVG_VAL",
command=lambda: controller.show_frame("PageFive"))
        btn2.pack()
        btn5.pack()


class PageOne(tkin.Frame):
def __init__(self, parent, controller):
        tkin.Frame.__init__(self, parent)
self.controller = controller
        label = tkin.Label(self, text="nerual", font=controller.title_font)
label.pack(side="top", fill="x", pady=10)
btn = tkin.Button(self, text="HOME",
command=lambda: controller.show_frame("StartPage"))
btn.pack()
```

```
        load = Image.open("./sad.png")
        render = ImageTk.PhotoImage(load)

image = tkin.Label(self, image=render)
image.image = render
image.place(x=0, y=0)

        load = Image.open("./neutral.PNG")
        load = load.resize((128, 128))
        render = ImageTk.PhotoImage(load)

        image = tkin.Label(self, image=render)
image.image = render
image.place(x=500, y=50)

class PageFive(tkin.Frame):

def __init__(self, parent, controller):
    tkin.Frame.__init__(self, parent)
self.controller = controller
    label = tkin.Label(self, text="MATRIX VALUE", font=controller.title_font)
label.pack(side="top", fill="x", pady=10)
btn = tkin.Button(self, text="HOME",
command=lambda: controller.disp_frame("StartPage"))
btn.pack()

    load = Image.open("./matrix.PNG")
    load = load.resize((428, 428))
    render = ImageTk.PhotoImage(load)

    image = tkin.Label(self, image=render)
image.image = render
image.place(x=100, y=100)
```

```
if __name__ == "__main__":
appl = MusicApp()
appl.mainloop()
```

PAPER PUBLICATION STATUS

Publication process not yet started

PLAGIARISM REPORT

Final report

ORIGINALITY REPORT

2%	0%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to VIT University Student Paper	1%
2	Submitted to St. Patrick's College Student Paper	<1%
3	Submitted to Coventry University Student Paper	<1%
4	Lecture Notes in Computer Science, 2006. Publication	<1%
5	Submitted to Chandigarh University Student Paper	<1%
6	Submitted to University of Applied Sciences Berlin Student Paper	<1%
7	Submitted to University of Technology, Sydney Student Paper	<1%
8	cs231n.stanford.edu Internet Source	<1%