A PROJECT REPORT ON

"Music Recommendation System Based on Facial Emotions"

Submitted

In the partial fulfilment of the requirements for

The award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

By

P. Akash Babu(171FA04231)
P. Yasaswi(171FA04229)
K. V. N. S. Chakradhar(171FA04279)

Under the esteemed guidance of

Mr. S. V. Rama Krishna, Asst. Professor

Dr. Dondeti Venkatesulu, Professor and HOD

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

(Accredited by NAAC "A" grade)

Vadlamudi, Guntur.

A PROJECT REPORT ON

"Music Recommendation System Based on Facial Emotions"

Submitted

In the partial fulfilment of the requirements for

The award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

By

P. Akash Babu(171FA04231)
P. Yasaswi(171FA04229)
K. V. N. S. Chakradhar(171FA04279)

Under the esteemed guidance of

Mr. S. V. Rama Krishna, Asst. Professor

Dr. Dondeti Venkatesulu, Professor and HOD

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

(Accredited by NAAC "A" grade)

Vadlamudi, Guntur.

VIGNAN'S FOUNDATION FOR SCIENCE TECHNOLOGY AND RESEARCH

(Accredited by NAAC "A" grade)

CERTIFICATE

This is to certify that the Project Report entitled **"Music Recommendation System based on Facial Emotions"** that is being submitted by **P. Akash Babu (171FA04231), K. V. N. S. Chakradhar (171FA04279), P. Yasaswi (171FA04229)** in partial fulfilment for the award of B.Tech degree in Computer Science and Engineering to the Vignan's Foundation for Science, Technology and Research, deemed to be University, is a record of bonafide work carried out by him under my supervision.

Mr. S. V. Rama Krishna **External Examiner** **Dr. Dondeti Venkatesulu**

Assistant Professor. **Professor, HOD**

ACKNOWLEDGEMENT

We are very grateful to our beloved Chairman **Dr. Lavu Rathaiah**, and Vice Chairman **Mr. Lavu Krishna Devarayalu,** for their love and care.

It is our pleasure to extend our sincere thanks to Vice-Chancellor **Dr. M.Y.S. Prasad** and Dean Engineering & Management, **Dr. M. Shanthi Sree Rukmini,** for providing an opportunity to do my academics in VFSTR.

It is a great pleasure for me to express my sincere thanks to **Dr. Dondeti Venkatesulu HOD, CSE** of VFSTR**,** for providing required infrastructure to do my Project.

We extend our whole hearted gratitude to **Mr. S. V. Rama Krishna**, under whose valuable guidance that the project came out successfully after each stage.

We feel it our responsibility to thank **Mrs. D. Radha Rani** and **Mr. Vijay Babu,** for helping in all required aspects during the course.

We thank all our **faculty** members and **Programmers** of Department of Computer Science and Engineering who provided technical support to us in our academics throughout course.

Finally, we wish to express thanks to our family members for the love and affection overseas and forbearance and cheerful depositions, which are vital for sustaining effort, required for completing this work.

With Sincere regards,

P. Akash Babu (171FA04231)
K. V. N. S. Chakradhar (171FA04279)
P. Yasaswi (171FA24229)

Date: 15-06-2021

ABSTRACT

Music plays a vital role in our everyday life. Life without music cannot be imagined. Music changes our mood; we listen to music when we are Happy. We also listen to music when we are Sad. We listen to music when we are bored. Music can induce a clear emotional response in its listeners. The pitch and rhythm of the music are managed in the areas of the brain that deal with emotions and mood. Thus, music plays an important role in enhancing our mood. As elders have said "Face is the Index of the Mind", the mood of a person can be known by looking at the face of the person. The abstract of this system/ project is to build an automated system that builds playlists and plays the songs according to the mood of the user by directly discerning the facial emotions of the user. This model requires a camera to capture the face of the user and then the mood of the user is recognized by CNNs. Then the playlist is recommended to the user based on the discerned "Mood" of the user. This eliminates the time-consuming and tedious task of manually segregating or grouping songs into different lists and helps in generating an appropriate playlist based on an individual's emotional features. Hence, the proposed system can be used to build a music recommendation system based on the facial emotion gestures of the user.

Keywords: Mood, Music, Emotion Recognition, Feature Extraction, DeepFace, Neural Networks.

Table of Contents

List of Figures & Tables

CHAPTER - 1

INTRODUCTION

Music induces a clear emotional response in its listeners[1]. Musical preferences have been demonstrated to be highly correlated with personality traits and moods[2]. Facial expressions are the most ancient and natural way of conveying emotions, moods and feelings[1]. As a deep learning neural network, convolutional neural network plays an extremely important role in face image recognition. Cognition technology of convolutional neural network and automatic music recommendation system is developed to identify a model that recognizes facial expressions and recommends music according to corresponding mood of the user.

Human beings have the natural ability to look at someone's face and guess their mood. This ability if learnt by an electronic device - computer, humanoid robot or a mobile device - can have valuable applications in the real world. Music, a tool for arousing emotions and feelings, is far more powerful than language[3]. Music is something which taps deeply into our emotional core as human beings. Thus, listening to good music can help us elevate our mood from a negative sense to a positive sense. For example, listening to upbeat songs when the person is feeling sad can help him come out of his sadness and start feeling better[4]. This system proposes one such application, emotion-based music recommendation. Emotion of the user can be easily guessed by looking at his/her face. For this purpose of face detection and emotion recognition, studying the features from his/her face is necessary.

Fig 1.1 Facial expression with different emotions [5]

The problems associated with face detection include background elements, lighting conditions, pose and facial expression. This domain of face detection and emotion detection is currently a very active area of research due to development of Virtual Reality and Augmented Reality. Real-time face detection and recognition systems have limited functionality due to the varying quality of images because of the problems associated like background, illumination, etc. Hence, research and development for solutions related to these problems is an ongoing work.

Using traditional music players, a user had to manually browse through his playlist and select songs that would soothe his mood and emotional experience[6]. This task was labour intensive and an individual often faced the dilemma of landing at an appropriate list of songs. Other systems which detect the mood of the user by using facial expression have their time and memory complexity relatively high and hence fail in achieving a real-time performance. Even if they recognize the mood of the user then their selection of songs for making a playlist is such that it 1will just pick songs reflecting the current mood of the user and will not try to enhance his mood in any way. So, if the user is sad, In the current systems, user is provided with a list of songs with sad emotion which can degrade his/her mood further and can lead to depression.

So, the system proposed will detect the emotion of the user from his facial expressions. It will then provide the user with a playlist of songs, listening to which the user will feel better.

Fig 1.2 Facial expression with 7 different emotions [7]

1.1 Basic Concepts

Convolutional Neural Network

A CNN receives an image as an input in the form of a 3D matrix. The first two dimensions corresponds to the width and height of the image in pixels while the third one corresponds to the RGB values of each pixel.

CNNs consist of the following sequential modules (each one may contain more than one layer)

- Convolution
- ReLu activation function
- Pooling
- Fully connected layers
- Output layer

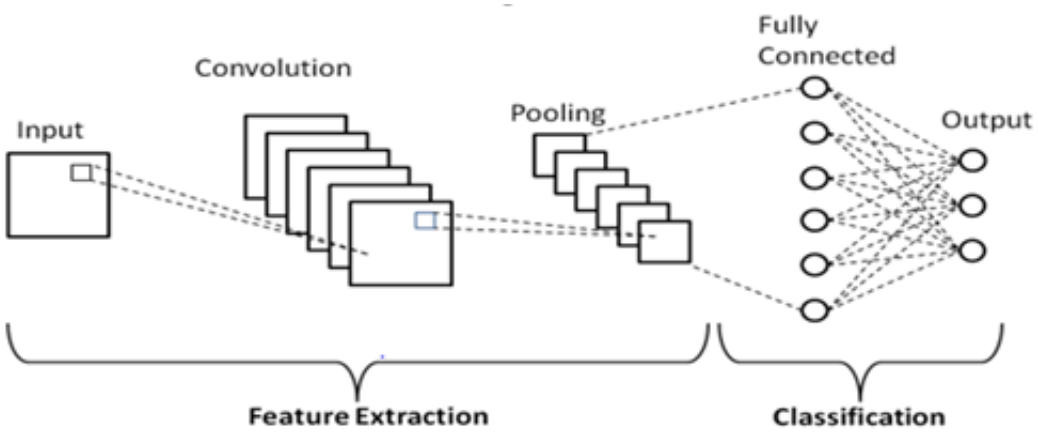


Fig 1.3 Schematic representation of Basic Convolutional Neural Network [8]

Convolution Layer :

The component associated with doing the convolution activity in the initial segment of a Convolutional Layer is known as the Kernel/Channel. Convolution activity is a component savvy network increase activity. Convolutional layers take the three-dimensional information framework and they pass a channel (otherwise called convolutional kernel) over the picture, applying it to a little window of pixels all at once (for example , 3x3 pixels) and moving this window until the whole picture has been filtered. The convolutional activity computes the dot result of the pixel esteems in the current channel window alongside the loads characterized in the channel. The yield of

this activity is the last tangled picture. The centre of picture order CNN's is that as the model trains what it truly does is that it learns the qualities for the channel grids that empower it to extricate significant highlights (shapes, surfaces, shaded regions, and so forth) in the picture. Each convolutional layer applies one new channel to the tangled picture of the past layer that can remove one more component. Thus, as we stack more channels, the more highlights the CNN can extricate from a picture.

Here are the three elements that enter into the convolution operation:

- Input image
- Feature detector
- Feature map

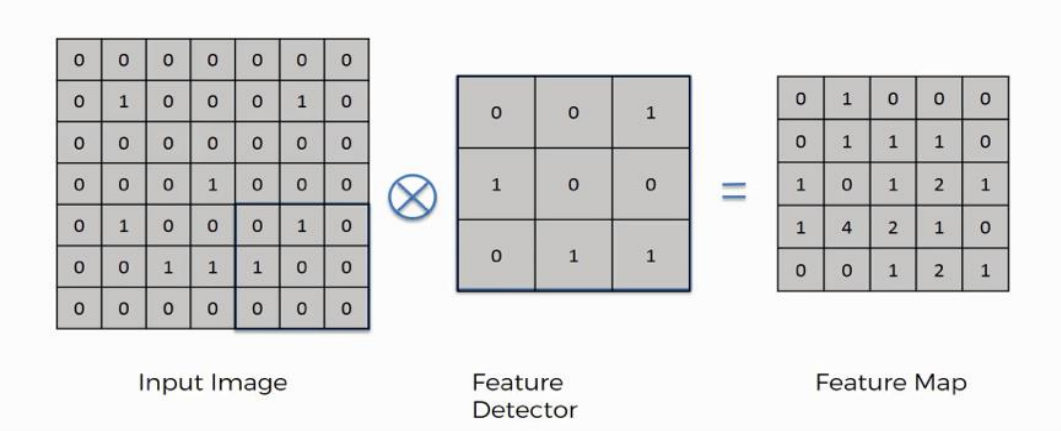


Fig 1.4 Feature Map generation through convolution operation [9]

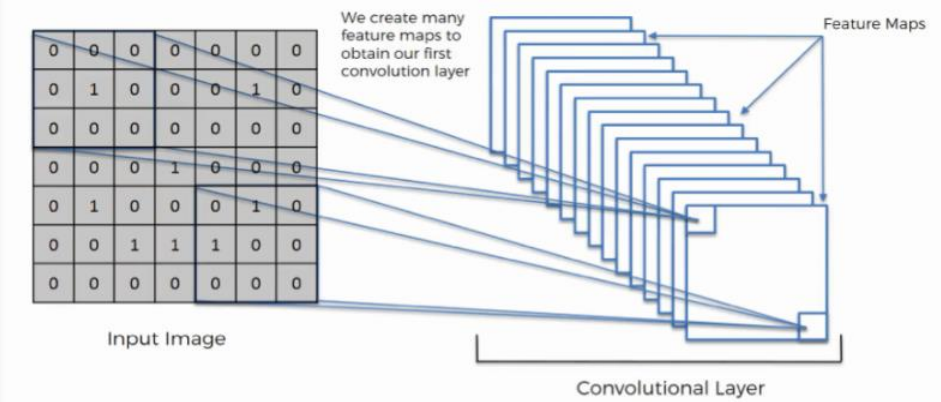


Fig 1.5 Creation of Convolution Layer [9]

ReLu Layer:

After every convolution activity, CNN applies to the yield a Rectified Linear Unit (ReLu) function to the convolved picture. If the convolved image has negative values, it replaces them with '0'.It also introduces nonlinearity into the model.

Pooling Layer:

Pooling is the process where dimension of the convolved image is reduced It does so to reduce processing time and the computing power needed. During this cycle, it saves the main feature data. There are a few techniques that can be utilized for pooling. The most widely recognized ones are Max pooling and Normal pooling. In our application, we will utilize max pooling as it is the best the greater part of the occasions. Max pooling is basically the same as the convolution interaction. A window slides over the feature map and concentrates tiles of a predefined size. For each tile, max pooling picks the most extreme worth and adds it to another feature map. Thus, the face features are extracted using convolution and pooling layers.

Fig 1.6 Maximum Pooled Feature Map [9]

Fully connected layer:

Fully connected layers are a fundamental part of Convolutional Neural Networks (CNNs), which have been demonstrated exceptionally fruitful in perceiving and classifying pictures for computer vision. The CNN interaction starts with convolution and pooling, separating the picture into features, and examining them freely. The consequence of this cycle takes care of into a fully connected neural network structure

that drives the last classification choice. In the Fully Connected Layer, all neurons of one layer are connected to all neurons in the next layer.

Output Layer:

The last fully connected layer is the output layer which applies a SoftMax function to the output of the previous fully connected layer and returns a probability for each class.

1.2 Motivation

- Music deftly plays with our emotions which in turn affect our mood. Music induces a clear emotional response in its listeners[10].
- Though, there are music recommendation systems based on happy and sad moods of the user still there are no effective algorithms and models which recognizes emotions like surprise , anger also effectively.
- Manual classification of songs based on mood, for making of a playlist, is time consuming and labour intensive. So , we thought of providing automatic music recommendation system based on human emotions [11]

1.3 Problem Statement

The task of manually Segregating or grouping the songs into different lists corresponding to different human emotions is time-consuming. So, this model detects the face in the image using OpenCV and extracts the facial expressions using convolutional neural network and then automatically recommend the songs corresponding to different human emotions.

1.4. Scope

- The project is limited to recognize the emotions Happy, Sad, Anger, Neutral and Fear.
- This system is limited to recommendation of Telugu songs.

1.5 Objective

- To predict human emotions based on facial expressions.
- To recommend songs based on the mood of the user.

1.6 Advantages

- Helps to identify live human emotions.
- Recommends music based on the emotion detected.
- Helps to make the user feel better by recommending happy & joyful songs when he is sad, calm& peaceful songs when he is angry.
- Avoids the risk of manually segregating the songs into different emotions because the platform recommends the songs automatically based on emotion detected.
- Provides a platform that can change the user mood into a good & relaxing mood by recommending songs to him/her

CHAPTER - 2

SYSTEM REQUIREMENTS SPECIFICATION

2.1 Hardware Requirements

We used a System with following configuration to conduct our proposed experiments.

Workstation	:	Dell
Processor	:	Intel Core
RAM	:	8 GB
GPU	:	2 GB (NVIDIA GEFORCE GTX)
Hard Disk Drive	:	512 GB SSD

2.2. Software Requirements

Operating System	:	Windows 10 Pro
Programming Language	:	Python
Deep Learning Framework	:	DeepFace
Supporting Libraries	:	NumPy, matplotlib, CV2, pygame

Tensor Flow – TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. Its flexible architecture allows for the easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices.

Keras – Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Features of Keras are:

- Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).
- Supports both convolutional networks and recurrent networks, as well as combinations of the two.
- Runs seamlessly on CPU and GPU.
- Keras is compatible with: **Python 2.7-3.6**

Pandas – Pandas is the most popular python library that is used for data analysis. It provides highly optimized performance with back-end source code is purely written in ***C*** or ***Python***.

NumPy – NumPy provides the essential multi-dimensional array-oriented computing functionalities designed for high-level mathematical functions and scientific computation.

Matplotlib – Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Fig. 2.1. Libraries used in this Project

CHAPTER - 3

ANALYSIS & DESIGN

3.1 Literature Survey

R.L.Rosa et al[12]. classified the music through sentences posted in various social media networks they did not classify through sentiments of users they classified through informal text which are posted in social media network where the text describes the emotions of a person whether he is happy, sad, or neutral through these sentence sentiment intensity analysis is done and using a new lexicon-based sentiment metric named it as "enhanced sentiment metric(ESM)".Sentiment analysis is a technique of natural language processing and text analysis these techniques can be applied to various fields like e-commerce, multimedia learning, etc, this analysis research is based on semantics, tag-based extractions, SVM. These techniques are used to describe emotions based on words and differentiating the emotions related to words. Here the sentiment analysis is divided into three. 1.lexicon-based 2. corpus-based 3. hybrid based. Lexicon based approach uses a word dictionary by using these words positive scale words and negative scale words are divided and these are represented using +, - symbols. The corpus-based approach uses a machine learning method where a large amount of data is collected

Xuan Zhu et al[13]. came up with an integrated music recommendation system technique that consists of automatic music genre classification, music emotion classification, and music similarity query with the help of AdaBoost algorithm, they proposed a new tempo feature called (LMFC) and it comes with timbre features which improve the performance of music classification here in this process where an interface is implemented where the user needs to press two buttons and the music will be played according to his genre selection of music there is also another feature where there will be user-selected songs where user can play a song on his selection the main goal of this process is to divide the large music databases into separate selections according to the specified genre.

Jukka Holm et al[14]. proposed a technique where music will be differentiated according to pictures here where the user interacts with pictures where they collect pictures according to emotions like happy, sad, etc where by using these pictures music will be

recommended where 40 Finnish people used this and it turns out to be 85% accuracy is showed on music selection there are few drawbacks where sometimes pictures selection will be failed and music recommendation to few pictures were also failed. Where they developed a hi-fi prototype application for a touch screen pc was developed and with the help of Nokia Research Centre's Super Music service music will be streamed.

Sanghoon jun et al[15]. focused on recognizing music emotion based on subjective human emotions they introduced three steps to achieve this. Different types of music features are extracted from the music signal. Those music features are observed and mapped into some values knows as (AV values) by a fuzzy inference engine. Finally, emotion will be shown based on AV values. Later introduced a feedback method where user can give feedback where they can rectify the flaws in the interface. There are few drawbacks where the user interface for expressing emotion is hard.

Ziyang Yu et al[16]. proposed that As there is the rapid growth of big data, deep learning with the help of these technologies where they combined micro-expression recognition technology of convolutional neural network and automatic music recommendation algorithm by using this technology where music will be suggested automatically by recognizing the emotion of a person there are few basic steps to achieve these firstly pictures need to be collected of all expressions later pre-processing of the images should be done and finally, classification should be done after implementing this there was 62.1% of recognition rate here the recognition will not be done by past listening history it will be on the user's present emotions after achieving this there are few problems to be solved sometimes the accuracy of the image processing will be slower and music recommendation algorithm should be improved for more accuracy.

Kyoungro Yoon et al[17]. suggested an autonomous and adaptive recommendation system that relies on the user's mood and implicit feedback to recommend songs without any prior knowledge about the user preferences. This method builds autonomously a latent factor model from the available online data of many users based on the associations extracted between the user, song, user mood, and song emotion. It uses a combination of the Reinforcement Learning framework and Page-Hinkley test to personalize the general song map for each mood according to user implicit reward.

Aurobind V.Iyer et al[11]. From the Emotion Based Mood Enhancing Music Recommendation system, stated the use of a different approach rather than asking the users mood from the application, the use of the photograph of the user which is sent through Fishers Face algorithm which recognizes the emotion of the user and then selects playlists from the servers.

Qing-Qiang Liu et al[18]. stated that the core of the recommendation framework is the construction of the product vs customer emotion model by two-dimensional overlap spaces, which plays an important role in conveying emotions in products. This method investigates the product feature extraction and proposes some related matching algorithms for the construction of the product vs customer emotion model.

Yu-Hao Chin et al[19]. pitched the idea of an emotion profile-based music recommendation system. Within the planned algorithmic rule, two feeling profiles area unit created exploitation call worth in support vector machine (SVM) and supported short term and future options severally. The recognized emotion, feeling profile, and private historical question results area unit fed into the recommended module to come up with the result music list.

Chaima Dhahri et al[20]. drafted an autonomous and adaptive recommendation system that relies on the user's mood and implicit feedback to recommend songs without any prior knowledge about the user preferences. Our method builds autonomously a latent factor model from the available online data of many users based on the associations extracted between the user, song, user mood, and song emotion. It uses a combination of the Reinforcement Learning framework and Page-Hinkley test to personalize the general song map for each mood according to user implicit reward.

3. 2. Existing Methodologies

3.2.1 Emotion Based Mood Enhancing Music Recommendation[11]

Methodology:

This model proposes a system ‘EmoPlayer’, an Android application, which help to minimize these efforts by suggesting the user a list of songs based on his current emotion. The system captures user’s image using camera and detects his face. It then detects the emotion and makes a list of songs which will enhance his mood as the songs keep playing.

EmoPlayer uses Viola Jones algorithm for face detection and Fisher faces classifier for emotion classification.

Advantages:

Images usually lie in high dimensional space which requires large memory for its storage and retrieval and also processing power. Also, the cost of computation on such high dimensional data is often too high . So, they used a technique to reduce the dimensionality of these data. Eigenface and Fisher faces are methods used for dimensionality reduction.

Disadvantages:

This model detects only happy and sad emotions. They need to develop an automatic music genre classification system and automatic music emotion classification system.

3.2.2 Facial Emotion Recognition Using Deep Convolutional Neural Network [21]

Methodology:

The main focus of this work is to create a Convolutional Neural Network (CNN) model that classifies different human facial emotions. The model is trained, tested and validated using the manually collected image dataset. Using python as the programming language, the model is implemented. The entire model is simulated in the Jupyter Notebook.

Advantages:

- The model classifies different facial emotions from the image dataset. The model has comparable training accuracy and validation accuracy which convey that the model is having a best fit and is generalized to the data.
- The model uses an Adam optimizer to reduce the loss function and it is tested to have an accuracy of 78.04%.

Disadvantages:

- The work can be extended to find out the changes in emotion using a video sequence which in turn can be used for different real time applications such as feedback analysis, etc.

3.2.3 A Music Recommendation Method with Emotion Recognition Using Ranked Attributes [22]

Methodology:

Here, they propose a new concept called Ranked Attributes that are useful to make reasonable music recommendations. More precisely, they propose to consider additional attributes to emotion, such as weather and time, and build a Ranked Attributes Tree

(RAT) that enables to recommend a music piece based on a combination of all ranked attributes. And also, they describe the following parts of the proposed method: database design, voice and emotion recognition, and music recommendation. A database design that contains of various tables to store keywords of emotion, weather and time, a voice recognition method, which receives the user's voice via smartphone microphone, and extracts text from user's voice using Google. Text-To-Speech API, an emotion recognition method that matches the emotion from the text and compares it to the keywords of an emotion in database.

Advantages:

Extremely fast feature computation, efficient feature selection, instead of scaling the image itself we scale the features.

Disadvantages:

Detector is mostly effective only on frontal images of faces, sensitive to lighting conditions.

3.2.4 Facial Micro-expression Recognition [16]

Methodology:

In Image Pre-processing, every image is converted to 48x48 pixel image and it is also converted to grayscale image. Thus, 2-Dimensional matrix is given as input to CNN. Micro expression detection and Feature extraction ,used an 8-layer CNN model with 4 convolution layers,2 pooling layers to improve the extraction capability of CNN network and finally 2 fully connected layers and the SoftMax classifier to map the extracted features into the classification probability. After every 2 convolution layers, there is one pooling layer and after every convolution layer ReLu Activation function is applied. The first fully connected layer has 512 neurons and the output layer has 7 neurons in it. After training the model, recognition rate is 62.1% . FER2013 dataset is used to train the model. It consists of 35,886 facial expression pictures and a total of 7 expressions labelled with 0 to 6. are there which are 0:Angry 1:Disgust 2:fear 3:Happy 4:Sad 5:Surprise 6:Neutral.

Advantages:

- Compared with the existing algorithms, the algorithm proposed in this paper increases the user's emotion recognition accuracy.
- It has the highest accuracy rate 83% on Happy and surprise tags.

Disadvantages:

- The model will have a deviation in identifying Disgust tags, and the probability of identifying as Angry is 51%.
- The probability of finding Disgust is only 29%; this model has the lowest recognition accuracy on Fear tags, only 23%.

2.5 Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation [10]

Methodology:

For pre-processing the image, they used OpenCV for face detection in the image before feeding the image into the layer. Pre-trained filters from Haar Cascades are used to quickly find and crop the face. The cropped face is then converted into grayscale and resized to 48*48 pixels. A multi-layered convolutional neural network is programmed to evaluate the features of the user image. This network architecture contains 9 convolutional layers with one max-pooling layer after every three convolutional layers followed by 2 dense layers. ReLu is applied after each convolutional layer. The final network was trained on 20,973 images and tested on 5,244 images. At the end, the model achieved an accuracy of 90.23%. Music Recommendation, they trained an artificial neural network which successfully classifies the songs in 4 classes with an accuracy of 97.69%. Dataset comprises of 390 songs spread across four moods. The distribution of the songs is as follows: class A with 100 songs, class B with 93 songs, class C with 100 songs and class D with 97 songs. The songs were manually labelled. Class A comprises of exciting and energetic songs, class B has happy and joyful songs, class C consists of sad and melancholy songs, and class D has calm and relaxed songs.

Advantages:

- It achieves high accuracy in the “angry” category, while also performing appreciably well in the “happy” and “calm” categories.
- It efficiently maps the user’s emotion to the correct song class with an overall accuracy of 97.69%.

Disadvantages:

- It would be interesting to analyse how the system performs when all seven basic emotions are taken into consideration.
- Additional songs from different languages and regions can also be added to make the recommendation system more robust.

3. 3. Software Development Life Cycle

In AGILE methodology is a practice that promotes **continuous iteration** of development and testing throughout the software development lifecycle of the project. In the Agile model, both development and testing activities are concurrent, unlike the Waterfall model.

The Agile software development methodology is one of the simplest and effective processes to turn a vision for a business need into software solutions. Agile is a term used to describe software development approaches that employ continual planning, learning, improvement, team collaboration, evolutionary development, and early delivery. It encourages flexible responses to change.

The agile software development emphasizes on four core values.

1. Individual and team interactions over processes and tools
2. Working software over comprehensive documentation
3. Customer collaboration over contract negotiation
4. Responding to change over following a plan

Fig 3.3.1: Agile model [23]

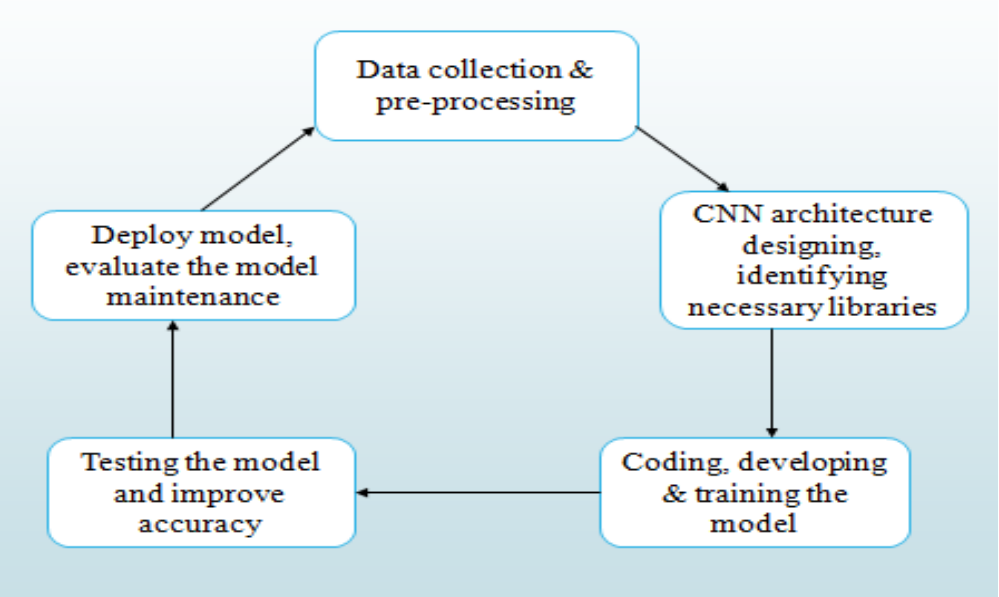


Fig 3.3.2 Agile software Development Lifecycle [23]

Agile projects require an iterative approach, which supports incremental, frequent, and consistent delivery of workable products to your customer or client. This innovative approach ensures your project team can consistently deliver concrete products without being delayed by changes and evolving requirements.

Agile has a high level of customer involvement and includes frequent reviews of progress with both the project team and the customer.

CHAPTER - 4

IMPLEMENTATION

4.1 Proposed System

The proposed system contains three modules namely Data Augmentation, Model Training& Testing, Face Detection & Emotion Recognition and Music Recommendation. The first two modules are done in the DeepFace library itself.

4.1.1 Process Flow Diagram

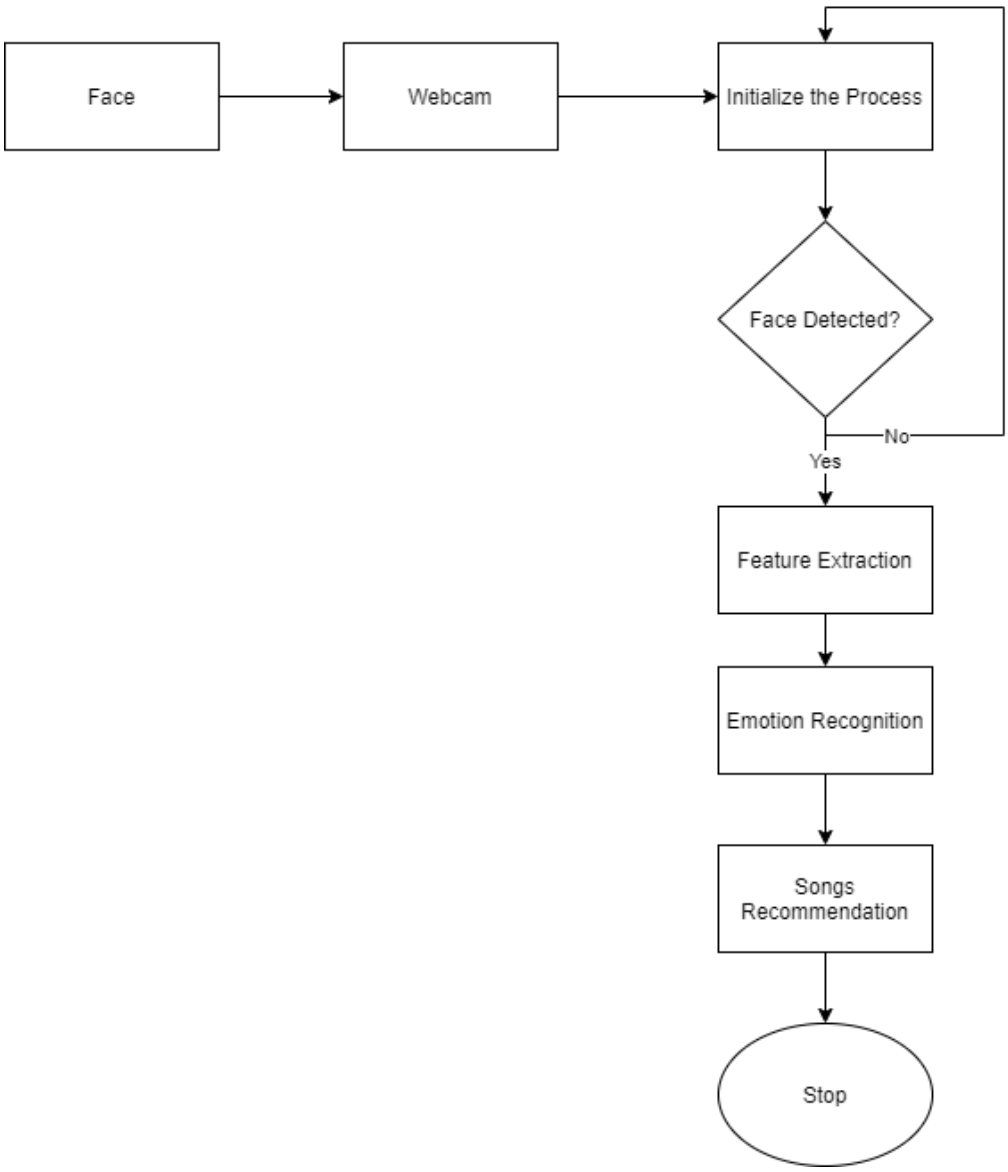


Fig 4.1.1: Flow oriented diagram for emotion detection music recommendation model

4.1.2. Use Case Diagram

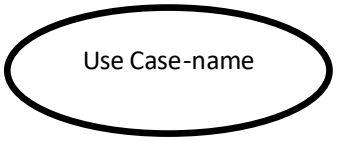
Use Case diagrams is used to identify the functionality provided by the system (use cases), the users who interact with the system (actors), and the association between the users and the functionality. Use Cases are used in the Analysis phase of software development to develop the high-level requirements of the system.

Use-case diagram notations:

The Following are the common notations used to represent use case diagram:

Use Case:

Use cases are used to represent high-level functionalities and it is used to know how the user will handle the system. A use case represents a distinct functionality of a system, a component, a relation, a package, or a class. The use case is denoted by an oval shape with the name of a use case written inside .The representation of a use case in UML is given below:



Actor:

It is used inside use case diagrams. The actor is an entity that interacts with the system to perform functionality. A human is the best example of an actor. An actor is an entity that initialize the use case from outside the scope of a use case. It can be any element, actor that can trigger an interaction with the use case. One actor can be related with multiple use cases in the system. The actor representation in UML is given below.

In use case diagram, the actors are User and Developer.

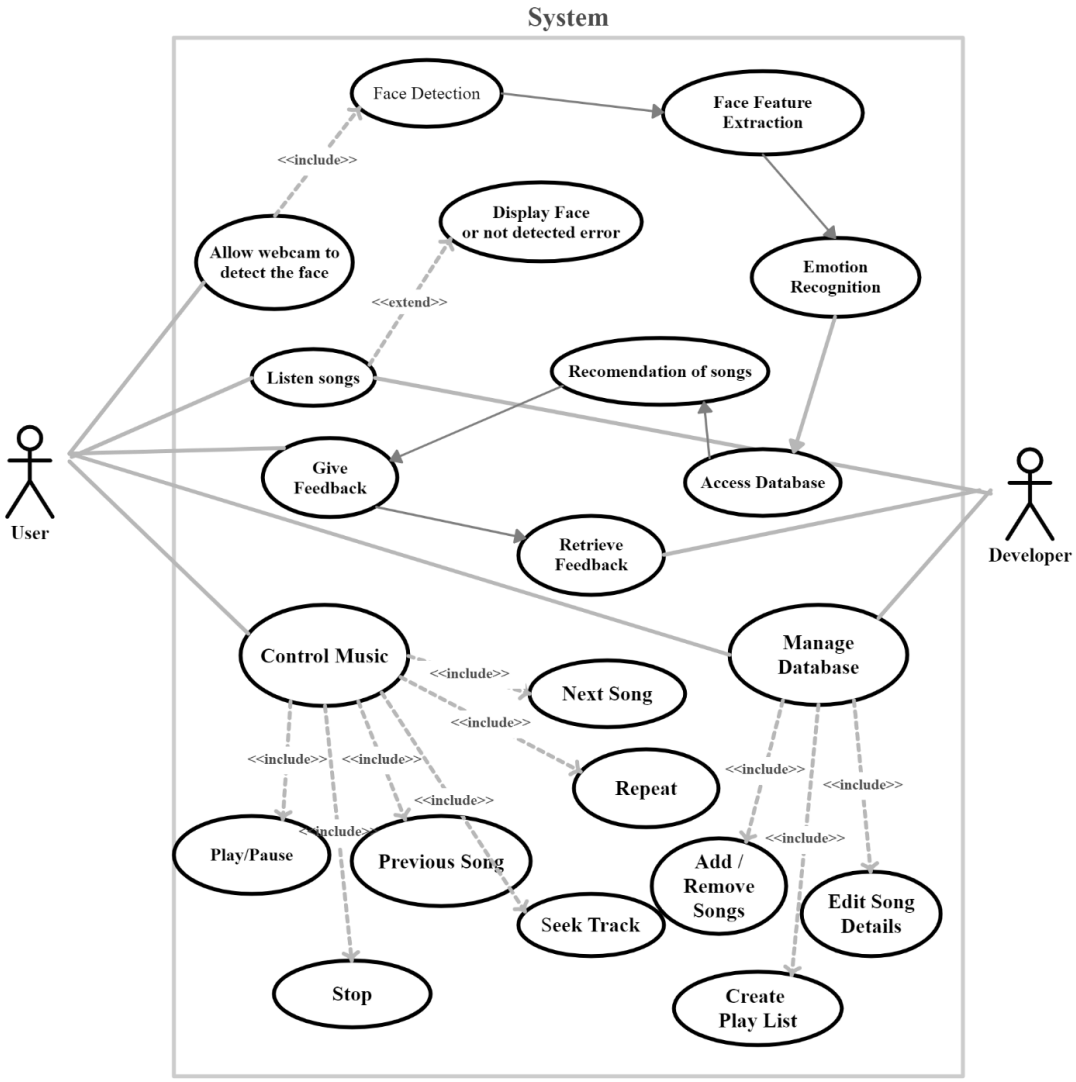


Fig 4.1.2: Use Case Diagram [23]

4.2 Methodology

Modules:

- Data Augmentation
- Model Training & Testing
- Face Detection & Emotion Recognition
- Music Recommendation

Data Augmentation:

- Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data.[24]
- The techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks.
- It acts as a regularizer and helps reduce overfitting when training a machine learning model.

Model Training & Testing:

DeepFace is the facial recognition framework utilized by Facebook for labelling pictures. It was created by scientists at Facebook AI Research (FAIR) at the 2014 IEEE Computer Vision and Pattern Recognition Conference (CVPR)[25].

In current DeepFace recognition there are 4 stages:

- Detect,
- Alignment,
- Represent,
- Classify

Lately, an enormous number of photographs have been slithered via web search tools, and transferred to informal organization by using this enormous data

The DeepFace device deals with a face check calculation, organized by man-made brainpower (AI) strategies utilizing neural organization models. A neural network is an assortment of prepared neurons that perform different actions and gives possible outcomes based on the emotion of a human being

As Facebook uses a high-level rendition of this methodology, the means are a smidgen more developed and explained than these. Adding the 3D change and piece-wise relative change in the system, the calculation is engaged for conveying more precise outcomes

The high level DeepFace calculation fills in as under

Working of DeepFace Recognition

Storing of the images to process

At the point when we upload pictures to the framework, the DeepFace tool checks the picture of the human face. Regardless of whether you are uploading two photographs of similar people, taken from various points, those are something similar for people yet not for the PC. It would be treated as two people by figuring calculations.

Matching

Then, at that point, it begins the 'matching' measure. This face distinguishing proof is done by creating and sending a sign to the Neural network, a nine-layer neuron network having a wide collection of human faces approx. 120 million.

Using of neural networking for facial feature detection

By accepting the signal, artificial neurons build up connection among synapses Every one of the associations between two neurons can be considered as a route. These routes pass the signal ahead with bit by bit to identify the uploaded image . These neurons check with all current pictures to pass the sign the wired way (picture is as of now existing or matches with any face) or terrible lighting and indicates as that picture isn't matching with some other faces. In the principal case, the signal proceeds with the route of the current face and stores the picture in the previous album via landmark testing, though in the subsequent case, The signal redirects its route on other outcomes like if the image consists of big eyes, it checks whether it had big eyes or small eyes. Assume the picture is having large eyes, the sign begins to go towards all huge eyes' pictures. Then, at that point further it can recognize whether the eyes are dark or earthy coloured. Assume it is dark, and afterward it can make further qualifications like the distance between eyes, shades of eyebrows, kinds of lips, nose, and so forth

68 Fiducial point testing

By following repeating identification routes using artificial intelligence, the DeepFace goes for 68 fiducial points testing, each human face is having 68 specific facial points that can make a match

Encoding and mapping

The tool checks pattern by coordinating with the testing image with different pictures of a same individual and a new individual. Then, at that point it encodes the right image. Later by encoding, it looks through the data like name and address of that encoding and saves the image in that collection.

Working of each stage in DeepFace Recognition

Initial stage is detect where it detects the image or face of the image to put in simple words where it captures the image and later alignment is done the main process happens in the alignment stage, Here the face I.e., detected will be transformed into 3D image, the main objective of this alignment is

The objective of this arrangement part is to produce frontal face from the input image that may contain faces from various posture and directions[25]. The technique proposed in this paper utilized 3D frontalization of faces dependent on the fiducial (face feature points) separate the frontal face. The entire process is shown below

Given an input image, we initially distinguish the face utilizing six fiducial points. These six fiducial points are 2 eyes, tip of the nose and 3 points on the lips[25]. These feature points are utilized to identify faces in the picture.

In this progression we create the 2D-face picture trimmed from the first picture utilizing 6 fiducial focuses.

In the third step, we apply the 67 fiducial point map with their relating Delaunay Triangulation on the 2D-aligned edited picture. This progression is done to adjust the out of plane turns. In this progression, we likewise create a 3D-model utilizing a generic 2D to 3D model generator and plot 67 fiducial points on that physically[25].

The last stage is frontalization of alignment. we add the residual part to x-y directions of 3D warp since it diminishes corruption in 3D-warp. At long last, frontalization is accomplished by doing piece-wise relative on Delaunay triangulation that we created on 67-fiducial focuses.

Any AI or deep learning framework needs sufficient training data so it can 'learn'. With a tremendous client base, Facebook has enough pictures to explore different avenues regarding. The group utilized in excess of 4 million facial images of in excess of 4000 individuals for this reason. This calculation plays out a great deal of activities for recognizing faces with human precision levels.

The end-product is a face representation which is produced from a 9-layer deep neural net. This neural net has in excess of 120 million factors, which are mapped to different locally connected layers. And the final result is Facebook can identify if the two pictures address a similar person. The site can do it, independent of environmental factors light, camera point, and shadings wearing on face for example facial make-up. Shockingly, this calculation works with 97.47 percent exactness, which is practically equivalent to natural eye precision 97.65 percent.

Face Detection& Emotion Recognition:

In this module, we use load model library form keras.models. We first capture the user face through web camera using cv2 library and read a frame from that video. Now we use Cascade Classifier to detect the face in the image(frame) detected and we crop the face and store it another object. We convert this image to grey color and resize this image to 48*48-pixel image as the model is trained on the same size of images which is 48*48.Now,we convert this image to array using img_to_array function imported from keras.preprocessing library. Now we use the load_model function to load the model that is already saved. Since we use 'SoftMax' function, it gives the probability for each class, so among them we take the maximum probability and we mark it as the emotion of the user. If the face is not found in the window, then "Face Not Detected' message appears. Thus, emotion of the user is identified.

Music Recommendation:

First, we create a folder named music system under this folder we create a folder called 'songs', under this folder again we create 5 more folders which corresponds to each of the 5 emotions. The folder names are Angry, Happy, Sad, Fear, Neutral. In each of these 5 folders, we play corresponding songs that match the user mood. In the code, first we need to import the necessary libraries tkinter, matplotlib, pygame. We recommend music to the users using tkinter. Tkinter is the standard GUI library for Python. Python when combined with tkinter provides a fast and easy way to create GUI applications[26]. We

first create an object for tkinter. Then, we define the size of the GUI window. We then store the necessary images in different objects by specifying their paths for later use. We create a class called 'player' in which we define 5 methods. First method is create frames() which creates 3 frames of different sizes as needed. Another function tract_widgets() is to display an image and for song tracking which displays the song name of the song that is currently being played.

Another function retrieve_songs() is to fetch the songs from the system according to the emotion detected and to display them on the frame right side.

Another function control_widgets() is to control the music player which can be play the music, pause the song, go to previous song, go to next song, increase or decrease volume.

This is how we created a music player using tkinter.

CHAPTER - 5

TESTING

The results, i.e., Output Screenshots of the proposed system under different testcases is as follows.

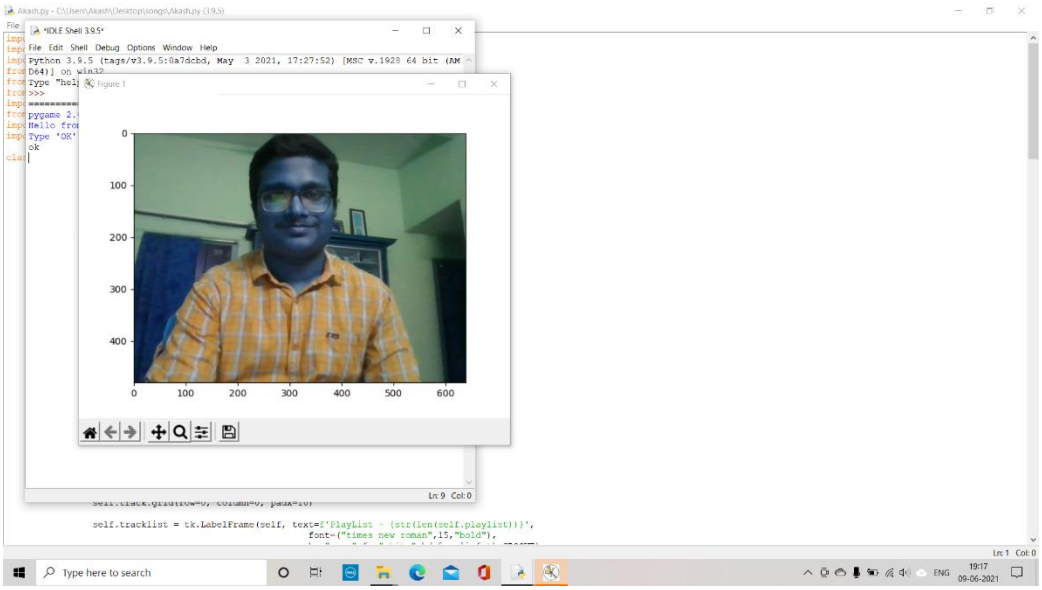


Fig. 5.1. Captured a Happy Face

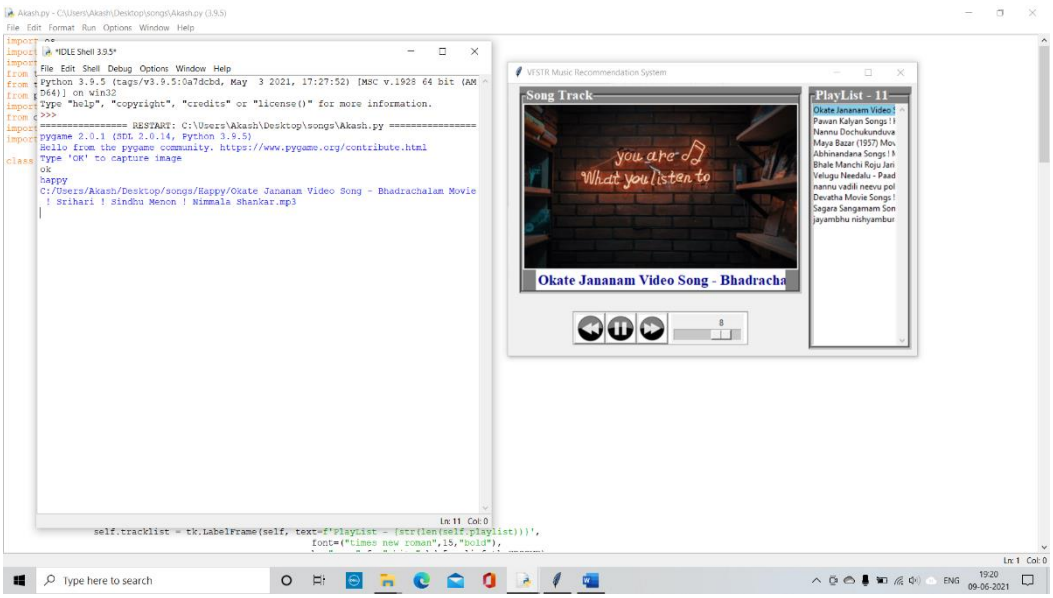


Fig. 5.2. Output for Happy Emotion

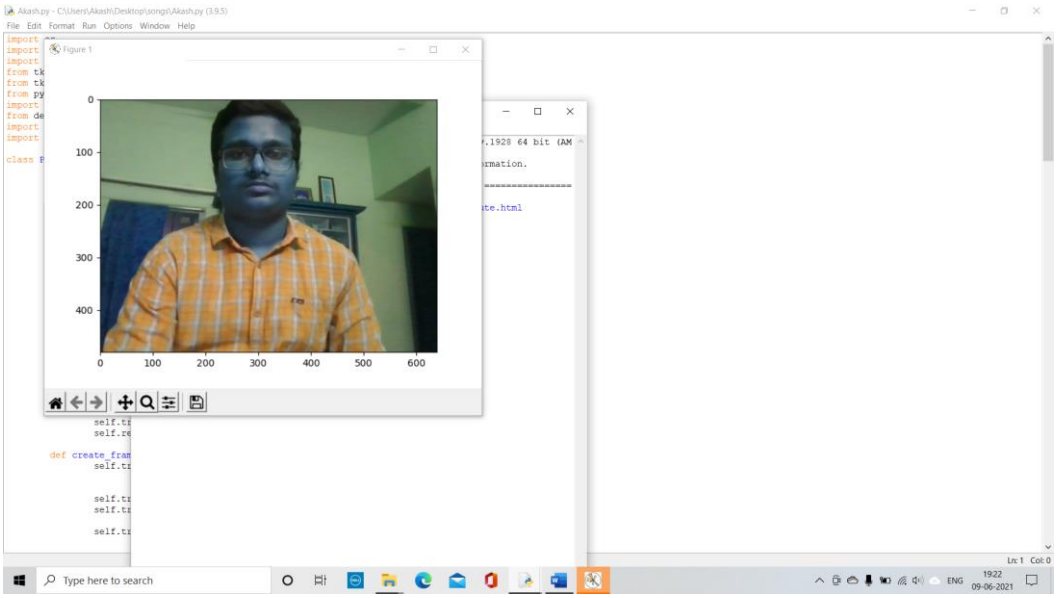


Fig. 5.3. Capturing a neutral Face

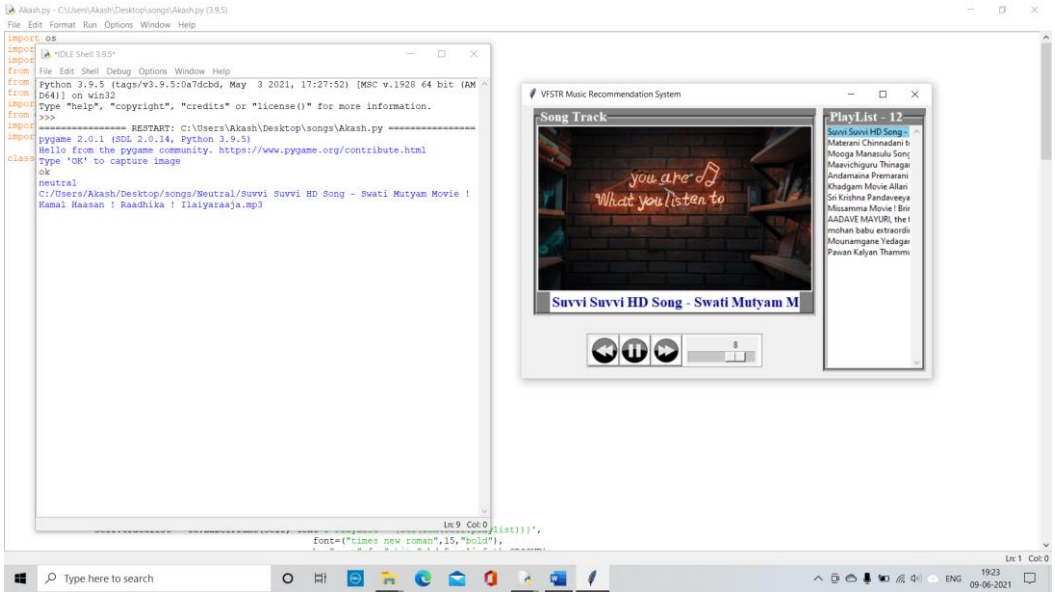


Fig. 5.4. Output for Neutral Emotion

CHAPTER – 6

CONCLUSION & FUTURE WORK

We would like to do much accurate model in the future and also a modem which detects all the basic 7 emotions which are happy, sad, anger, neutral, fear, disgust, surprise accurately. In the future we want to develop an android app for emotion-based music recommendation system which can be installed in our Android phones easily. We also plan to develop some special features where quotation of some great personalities will be recommended based on the user emotions if the user is detected as sad a song was suggested according to that song a quote will be displayed so that the user can be motivated. People tend to listen lyrics a lot at the same time if quotation were showed they tend to feel motivated or feel active and where they can share it in social media this help at least few people to motivate. And other feature is that if user uses the application most then some points will be added and these points can be used to purchase singer albums. One more thing we want to include is we wish to recommend songs of different languages like Hindi, English, Tamil etc.

REFERENCES

[1] "Smart Music Player Integrating Facial Emotion Recognition." https://www.ukessays.com/essays/computer-science/smart-music-player-integrating-facial-emotion-9441.php (accessed Jun. 09, 2021).

[2] "(No Title)." http://www.adalyajournal.com/gallery/istj-192.pdf (accessed Jun. 09, 2021).

[3] "Music & How It Impacts Your Brain, Emotions." https://psychcentral.com/lib/music-how-it-impacts-your-brain-emotions#1 (accessed Jun. 09, 2021).

[4] B. S. Chhabra, "Emophony-Face Emotion Based Music Player," *Int. Res. J. Eng. Technol.*, 2020, Accessed: Jun. 09, 2021. [Online]. Available: www.irjet.net.

[5] "RAVDESS Emotional speech audio | Kaggle." https://www.kaggle.com/uwrfkaggler/ravdess-emotional-speech-audio (accessed Jun. 08, 2021).

[6] "(No Title)." https://www.irjet.net/archives/V5/i4/IRJET-V5I41065.pdf (accessed Jun. 09, 2021).

[7] "Smart Doll - Development Eyes of Things." http://eyesofthings.eu/?page_id=1884 (accessed Jun. 08, 2021).

[8] "Schematic diagram of a basic convolutional neural network (CNN)... | Download Scientific Diagram." https://www.researchgate.net/figure/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26_fig1_336805909 (accessed Jun. 08, 2021).

[9] "The Ultimate Guide to Convolutional Neural Networks (CNN) - Blogs SuperDataScience - Machine Learning | AI | Data Science Career | Analytics | Success." https://www.superdatascience.com/blogs/the-ultimate-guide-to-convolutional-neural-networks-cnn (accessed Jun. 08, 2021).

[10] S. Gilda, H. Zafar, C. Soni, and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," *Proc. 2017 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2017*, vol. 2018-Janua, pp. 154–158, 2018, doi: 10.1109/WiSPNET.8299738.

[11] A. V. Iyer, V. Pasad, S. R. Sankhe, and K. Prajapati, "Emotion based mood enhancing music recommendation," *RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc.*, vol. 2018-Janua, pp. 1573–1577, 2017, doi: 10.1109/RTEICT.2017.8256863.

[12] R. L. Rosa, D. Z. Rodríguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359–367, 2015, doi: 10.1109/TCE.2015.7298296.

[13] X. Zhu, Y. Y. Shi, H. G. Kim, and K. W. Eom, "An integrated music recommendation system," *IEEE Trans. Consum. Electron.*, vol. 52, no. 3, pp. 917–925, 2006, doi: 10.1109/TCE.2006.1706489.

[14] A. Lehtiniemi and J. Holm, "Using animated mood pictures in music recommendation," *Proc. Int. Conf. Inf. Vis.*, pp. 143–150, 2012, doi: 10.1109/IV.2012.34.

[15] S. Jun, S. Rho, B. J. Han, and E. Hwang, "A fuzzy inference-based music emotion recognition system," *IET Conf. Publ.*, no. 543 CP, pp. 673–677, 2008, doi: 10.1049/cp:20080398.

[16] Z. Yu, M. Zhao, Y. Wu, P. Liu, and H. Chen, "Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition," *Chinese Control Conf. CCC*, vol. 2020-July, pp. 7257–7263, 2020, doi: 10.23919/CCC50068.2020.9189600.

[17] K. Yoon, J. Lee, and M. U. Kim, "Music recommendation system using emotion triggering low-level features," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 612–618, 2012, doi: 10.1109/TCE.2012.6227467.

[18] Q. Q. Liu, J. Zhu, and T. Kong, "RSED: A novel recommendation based on emotion recognition methods," *Proc. - 2009 Int. Conf. Inf. Eng. Comput. Sci. ICIECS 2009*, no. 2, pp. 1–4, 2009, doi: 10.1109/ICIECS.2009.5364111.

[19] Y. H. Chin, S. H. Lin, C. H. Lin, E. Siahaan, A. Frisky, and J. C. Wang, "Emotion profile-based music recommendation," *Proc. - 2014 7th Int. Conf. Ubi-Media Comput. Work. U-MEDIA 2014*, pp. 111–114, 2014, doi: 10.1109/U-MEDIA.2014.32.

[20] C. Dhahri, K. Matsumoto, and K. Hoashi, "Mood-aware music recommendation via adaptive song embedding," *Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018*, no. 2, pp. 135–138, 2018, doi: 10.1109/ASONAM.2018.8508569.

[21] E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 317–320, 2020, doi: 10.1109/ICACCS48705.2020.9074302.

[22] S. H. Park, S. Y. Ihm, W. I. Jang, A. Nasridinov, and Y. H. Park, "A music recommendation method with emotion recognition using ranked attributes," *Lect. Notes Electr. Eng.*, vol. 330, pp. 1065–1070, 2015, doi: 10.1007/978-3-662-45402-2_151.

[23] "Software Engineering: A Practitioner's Approach." Accessed: Jun. 08, 2021. [Online]. Available: www.mhhe.com/pressman.

[24] "Urwa Muaz – Medium." https://urwamuaz.medium.com/ (accessed Jun. 09, 2021).

[25] “Deep Face Recognition - GeeksforGeeks.” https://www.geeksforgeeks.org/deep-face-recognition/ (accessed Jun. 09, 2021).
[26] “Python - GUI Programming (Tkinter) - Tutorialspoint.” https://www.tutorialspoint.com/python/python_gui_programming.htm (accessed Jun. 09, 2021).