“EBASS - Emotion Based Assistant Service System”

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENT FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By

ABHIJIT DEURI (16/376)

HIRAK JYOTI BUNGRUNG (16/378)

RISHIK CHANDAN (17/548)

Guided By

MRIDUL JYOTI ROY

Assistant Professor

2016-20

GAUHATI UNIVERSITY, GUWAHATI

ASSAM ENGINEERING COLLEGE, JALUKBARI

GUWAHATI-781013

July-2020

“EBASS - Emotion Based Assistant Service System”

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENT FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By

ABHIJIT DEURI (16/376)

HIRAK JYOTI BUNGRUNG (16/378)

RISHIK CHANDAN (17/548)

Guided By

MRIDUL JYOTI ROY

Assistant Professor

2016-20

GAUHATI UNIVERSITY, GUWAHATI

ASSAM ENGINEERING COLLEGE, JALUKBARI

GUWAHATI-781013

July-2020

CONTENTS

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ASSAM ENGINEERING COLLEGE::JALUKBARI
GUWAHATI-781013

Forwarding Certificate

This is to certify that **ABHIJIT DEURI (16/376), HIRAK JYOTI BUNGRUNG (16/378), RISHIK CHANDAN (17/548)** has/have carried out the project work **EBASS (Emotion Based Assistant Service System)** under the supervision of **Mridul Jyoti Roy** and has/have compiled this thesis reflecting the candidate's work in the semester long project. The candidate(s) did this project full time during the whole semester and the analysis, results, claims etc. are all related to his/her/their studies/study and works during the semester.

We recommend submission of this thesis as the partial fulfillment of the requirement for the degree of Bachelor of Engineering in Computer Science & Engineering of Gauhati University.

Prof. Dinesh Shankar Pegu
Head of the Department (HOD)
Computer Science & Engineering
Assam Engineering College
Jalukbari, Guwahati – 781013

External

Name: …………..……...
Signature:…...……..……
Affiliation: ………...…..

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ASSAM ENGINEERING COLLEGE::JALUKBARI

GUWAHATI-781013

Forwarding Certificate

This is to certify that **ABHIJIT DEURI (16/376), HIRAK JYOTI BUNGRUNG (16/378), RISHIK CHANDAN (17/548)** has/have carried out the project work **EBASS (Emotion Based Assistant Service System)** under my supervision and has/have compiled this thesis reflecting the candidate's work in the semester long project. The candidate(s) did this project full time during the whole semester and the analysis, results, claims etc. are all related to his/her/their studies/study and works during the semester.

I recommend submission of this thesis as the partial fulfillment of the requirement for the degree of Bachelor of Engineering in Computer Science & Engineering of Gauhati University.

Mr. Mridul J. Roy

(Assistant Professor)

Signature:

Department of

Computer Science & Engineering

Assam Engineering College, Jalukbari

Guwahati - 781013

ACKNOWLEDGEMENTS

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. All that we have done is only due to such supervision and assistance and we would not forget to thank them. We would like to express our heartfelt gratitude to each and every person connected with the completion of this project. We would like to thank God for his eternal presence.

We are grateful to our guide Mridul Jyoti Roy, Asst. Prof., Department of Computer Science and Engineering, Assam Engineering College, for guiding us throughout the course of our work and providing significant input for the completion of our project. Secondly, we would like to thank Mr Dinesh Shankar Pegu, HOD, Department of Computer Science and Engineering, and Dr. Atul Bora, Principal, Assam Engineering College, Jalukbari, Guwahati-13, for giving us the opportunity to carry out our research project work in this prestigious Institution. We would also like to take this opportunity to thank our family and friends who provided us continuous mental support which helped us in the completion of this project.

Assam Engineering College

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GAUHATI UNIVERSITY

Declaration by the Candidate

We **ABHIJIT DEURI (16/376), HIRAK JYOTI BUNGRUNG (16/378), RISHIK CHANDAN (17/548)** BE. student(s) of the Department of Computer Science & Engineering, Assam Engineering College hereby declares that I/we have compiled this thesis reflecting all my/our works during the semester long full time project as a part of my BE curriculum.

We declare that we have included the descriptions etc of my project work, and nothing has been copied/replicated from other’s work. The facts, figures, analysis, results, claims etc depicted in my/our report are all related to my full time project work.

We also declare that the same report or any substantial portion of this project report has not been submitted anywhere else as part of any requirements for any degree/diploma etc.

Date:

16/376 Abhijit Deuri Signature…………….

16/378 Hirak Jyoti Bungrung Signature…………….

17/548 Rishik Chandan Signature…………….

ABSTRACT

People tend to increasingly accumulate stress and anger because of their daily lives. Certain activities such as – Listening to music, watching a video, assisting them for workouts, telling them a joke, etc assists them to reduce stress and anger. However, it may be unhelpful if the activity does not suit the emotion/mood of the person. Moreover, there is no assistance system that is able to suggest activities based on the user's emotion in order to help reduce stress and anger. To solve this problem, we propose a system that tries to change the mood of the user.

We aim to accomplish this by first detecting the user's emotion via a few parameters such as – Haar-Cascade to extract the **Facial image** and CNN to classify facial expression images into emotions. After detecting the said emotion, the system compiles a list of activities that may suit the person's mood and interacts with the user using **voice**, i.e – If the person's mood is detected as "sad", the system tries to make him happier. If the person's mood is detected as "angry", their **heart rate** is checked using Eulerian Video Magnification and then the system compiles a different list of activities that may calm the user. The system, thereby, has health benefits too.

LIST OF FIGURES

Chapter 1 - Introduction

1.1 Motivation

Mental health is a major concern worldwide and India is not far behind in sharing this. If we evaluate developments in the field of mental health, the pace appears to be slow. Dr. Brock Chisholm, the first Director-General of the World Health Organization (WHO), in 1954, had presciently declared that “without mental health there can be no true physical health.” More than 60 years later, the scenario has not altered substantially. About 14% of the global burden of disease is attributed to neuropsychiatric disorders. The burden of mental disorders is likely to have been underestimated because of inadequate appreciation of the inter-play between mental illness and other health disorders. There remain considerable issues of priority-setting based on the burden of health problems and of addressing inequalities in relation to determinants and solutions for health problems.

Progress in mental health service delivery has been slow in most low- and middle-income countries. Barriers include the existing public-health priorities and its influence on funding; challenges to delivery of mental health care in primary-care settings; the low numbers of those trained in mental health care; and the lack of mental health perspective in public-health leadership. There have been numerous calls for invoking political will, for enhancing advocacy and for galvanizing community participation; all with scant improvement in outcomes.

Thus, it becomes now opportune to explore the paradigm of mental health awareness as a means of combating stigma, enhancing prevention, ensuring early recognition, and also stimulating simple and practical interventions within the community. Today there are opportunities in terms of growing acknowledgement of mental disorders as key targets of global health action, as well as of leveraging new technologies particularly the internet, big data and cell phones in amplifying simple field interventions found successful in primary care and other echelons. [1]

The major motive to develop this project is the need to have an effective way to reduce negative emotions among people. Negative emotions in turn lead to deteriorating performance in day to day lives. People tend to increasingly accumulate stress and anger because of their daily lives. Certain activities such as – Listening to music, watching a video, assisting them for workouts, telling them a joke, etc helps them to reduce stress, anger and elevate sadness (which are considered negative emotions). Moreover, there is no assistance system that is able to suggest activities based on the user’s emotion in order to help reduce stress and anger. To solve this problem, we propose a system that tries to enhance a person’s mood from a negative emotion to a positive emotion and also suggests activities for the positive emotions as well.

1.2 Problem Statement

Negative emotions can lead to physical symptoms including headaches, upset stomach, elevated blood pressure, chest pain, and sleeping problems.

The Journal of the National Medical Association added that people who respond negatively to anger are 9% more likely to have heart attacks. While people who are constantly sad or clinically depressed can be prone to memory loss. [2]

Emotional control is a habit for some people. For others, emotional response is automatic.

Symptoms associated with being unable to control emotions include:

- being overwhelmed by feelings
- feeling afraid to express emotions
- feeling angry, but not knowing why
- feeling out of control
- having difficulty understanding why you feel the way you do
- misusing drugs or alcohol to hide or “numb” your emotions

Following are the effects of negative emotions -

1. Physical health

Constantly operating at high levels of stress and anger makes a person more susceptible to heart disease, diabetes, a weakened immune system, insomnia, and high blood pressure.

2. Mental health

Chronic anger consumes huge amounts of mental energy, and clouds a person's thinking, making it harder to concentrate or enjoy life. It can also lead to stress, depression, and other mental health problems.

3. Career

Constructive criticism, creative differences, and heated debate can be healthy. But lashing out only alienates one from their colleagues, supervisors, or clients and erodes their respect towards the person.

4. Relationships

Anger can cause lasting scars among people and can get in the way of family, friendships and work relationships. Explosive anger makes it hard for others to trust that person, speak honestly, or feel comfortable—and is especially harmful to children.

Negative emotions like fear, sadness, and anger are a basic part of life and sometimes we struggle with how to deal with them effectively. It can be tempting to act on what you're feeling right away, but that often doesn't fix the situation that caused the emotions. In fact, it may lead to more problems to deal with down the road. [3]

Some of the harmful ways that people deal with negative emotions:

Denial

Denial is when a person refuses to accept that anything is wrong or that help may be needed. When people deny that they are having problematic feelings, those feelings can bottle-up to a point that a person ends up “exploding” or acting out in a harmful way.

Withdrawal

Withdrawal is when a person doesn’t want to be around, or participate in activities with other people. This is different than wanting to be alone from time to time, and can be a warning sign of depression. Some people may withdraw because being around others takes too much energy, or they feel overwhelmed. Others may withdraw because they don’t think other people like them or want them to be around. In some cases, people who have behaviors that they are ashamed of may withdraw so other people don’t find out about what they are doing. But withdrawal brings its own problems: extreme loneliness, misunderstanding, anger, and distorted thinking. We need to interact with other people to keep us balanced.

Bullying

Bullying is when a person uses force, threats, or ridicule to show power over others. People typically take part in bullying behavior because they don’t feel good about themselves and making someone else feel bad makes them feel better about themselves or feel less alone. It is harmful to both the bully and the person being bullied and does not address underlying issues.

Self-Harm

Self-harm can take many forms including: cutting, starving oneself, binging and purging, or participating in dangerous behavior. Many people self-harm because they feel like it gives them control over emotional pain. While self-harming may bring temporary relief, these behaviors can become addictive and can lead people to be more out of control and in greater pain than ever.

Substance Use

Substance use is the use of alcohol and other drugs to make a person feel better or numb about painful situations. Alcohol and drug use can damage the brain, making it need higher amounts of substances to get the same effect. This can make difficult feelings even worse and in some cases, leads to suicidal thoughts or addiction. If you are concerned about your own or someone else’s use of drugs or alcohol, talk to a responsible adult right away to get help.[4]

Hence, it will be beneficial for all if our proposed system can help reduce the user’s negative emotions to some extent.

1.3 Proposed solution

The solution that we propose is using our system to detect the emotion of the user using Haar Cascade to detect the face region from a live video feed and then classify it into one of 4 emotions, namely “Happy”, “Neutral”, "Sad” and “Angry” using Convolutional Neural Network(CNN). After the emotion is detected, we try to suggest some activities based on the detected emotion. The activities for each emotion differ from each other and some activities take priority for some particular emotions such as “getting a movie recommendation” is prioritized for “Happy” and “Neutral” emotions but not for “Sad” and “Angry” emotions. Similarly, we do not suggest activities such as “Reading News on a particular topic” for the “Sad” emotion.

This is all done using our voice based assistant that interacts with the user via audio input/output to make the user feel more comfortable talking to the system rather than having to select an item/click an item while keeping in mind that this also helps blind people to interact with our system better. An angry person can also find that his/her heartbeat is checked prior to activities that help reduce anger so that an angry person can trust the system that his/her heartbeat has slowed down and they no longer feel uneasy.

1.4 Goals and Objectives

- To develop an effective assistant service system based on emotions of the user.

- To provide anger management services that in turn can help the individual develop a stronger sense of empathy, so they can better understand others perspectives.

- To help the user to keep track of the current emotion.

- To provide certain appropriate activities based on the user’s current emotion which if detected to be positive, the system will suggest activities to help the user maintain that emotion whereas if the emotion is detected to be negative then the system tries to elevate the user’s emotion and in turn change it from negative to positive.

- To implement the system as a voice assistant so that visually impaired people can access the system as well.

- To build the system in a way so that it can act as a direct by-product of stress and anger management therapy.

1.5 Specific objectives.

i. To design and implement a facial recognition system for capturing facial image input along with the emotion shown by the user.

ii. Designing a database that keeps a list of activities that the system will curate from depending upon the user’s emotion.

iii. Designing a system that takes the emotion recognized by our model and determines which set of activities are to be chosen.

iv. To design a heart rate measurement system that can detect the heart rate prior to a calming activity and then detect the heart rate again to check if there is a difference. [If there is a decrease in heart rate then the user has been calmed successfully to some extent.]

v. To integrate an Audio Input-Output system so that the user can feel comfortable saying his/her command into our system while receiving on screen text in voice.

vi. To implement hardware and software parts of the proposed system to reflect the real design process and integrate the software and hardware parts to make a complete system.

1.6 Significance of the Project.

The successful completion of the project and its implementation will have the following advantages;

- Suggestion of suitable activities based on the user’s current mood.
- Help to turn negative emotions into positive emotions.
- Help reduce suicidal thoughts by distracting the user with funny videos, jokes and other such activities when the emotion is detected as “Sad”
- Playing songs that are suitable to the user’s emotion
- Getting recommendations of the movies based on the user's likings and preferences.
- Good outlet for relieving stress and anger issues
- Activities to keep positive emotions intact.
- Assist blind people and serve as a good place to interact where they might feel socially difficult to do so.

1.7 Scope and Limitation of the Project.

Scope

- The project concerns with the detection of emotion and getting activities based off of the detected emotion,
- The whole system can be designed for a single particular user and based on his likings and preferences, the activities can be prioritized. This even includes the system “learning” about the user’s dislikes and similar disliked activities can be neglected.
- The system can also be able to classify the emotion based on the detected heartbeat of the user based on the factors that the heart rate depends upon such as Body to Mass Index (BMI), Age, Gender, etc.
- There is also the opportunity to add a module such that a live video feed of the room or house can be shown to the user where their pet or child is displayed on screen. This may prove to be a real mood lifting activity.
- The system could be made to keep track of the past emotions and also track the activities that had helped the user to enhance his mood
- Inclusion of a timer, To-Do List or smart services such as weather reports can make this system feel closer to being a virtual assistant.
- Detect the emotion of the user while it runs in background to keep track of the user’s emotion while he/she performs their daily activities.

Limitations

- Some people are not as expressive as other people and this group of people may find it hard to use our system as Facial Expression is really important to the detection of the emotion accurately.
- Proper lighting for the detection of the emotion is required. Therefore, usage of this system at low lighting conditions may introduce noise in the video feed, which is not desirable.

- While measuring the heart of the user, in order to get more accurate reading the extraction region of the facial input should be steady therefore the user’s head must be kept fixed and the camera must be placed at a particular distance.
- There should be a proper microphone attached to the system so that our Natural Language Processor (NLP) can understand what the user is saying
- Offline usage of this system may cause certain activities to not function properly. i.e- A system without proper internet connection may not be able to utilize our system to its full potential.

Chapter 2 - Literature Review

2.1 Introduction

Literature review involves taking time to read different documents from text of scholars offline and on web sites. Furthermore, it provides the necessary knowledge and information obtained from various information sources. The reading built up the knowledge and new techniques toward implementation of this project. These readings include substantive findings, theoretical and methodological contributions to a particular topic, consulting project supervisor, lecturers and other professionals to get a clear knowledge of the system to be implemented.

For the purpose of monitoring the heart rate of an individual, it is not always necessary to use wrist or chest worn or any other wearable device or piece of apparatus. The MIT Computer Science and Artificial Intelligence Laboratory (CSail) research group's public made algorithm can also be used which converts a video file to a file with certain specifiable frequencies of movement and/or colors amplified. This amplification can be used for visualizing and monitoring heart rate for drivers in order to sense whenever the driver develops any sense of sleepiness. A Kinect camera which is a motion sensing input device that can be mounted on the dashboard of the car which can be used to collect the video data input from which the driver's heart rate can be extracted in real time. This technology can be used to unobtrusively monitor heart rate and breathing of infants, sleep apnea patients or other individuals where worn devices are not optimal. [5]

A contact-less heart rate measurement technique is developed using video taken by a camera so that users can measure the heart rate without any medical devices. Attention is focused on the blood oxygenated haemoglobin which absorbs green light and measures the heart rate by detecting the variation of the intensity of green pixels in a person's facial image. In previous researches, it wasn't possible to measure the heart rate with precision when a person was moving. But in this proposed method, the heart rate can be accurately measured even while moving the face. The results showed a performance of detection error smaller than 3 bpm for distances up to 6.0 m and lightning conditions over 500 lux. [6]

Nowadays, facial recognition is implemented in security systems to grant access to areas that are only allowed for authorized people. However, an additional layer of security can be added to these systems by determining if the person in front of the camera is present in actual time and that the detected object is not a 2D representation of that individual i.e a photograph. The paper focuses on real-time emotion detection. Therefore, a novel algorithm is developed to extract emotions based on the movement of 19 feature points. These feature points are located in different regions of the face such as the eyebrows, nose, mouth and eyes. To obtain the feature points, an Ensemble of Regression Trees is constructed. After the extraction of the feature points; 12 distances, in and around these facial regions, are calculated to be used in displacement ratios. In the last step, the algorithm inputs the displacement ratios to a classification algorithm, which is a cascade of a multi-class support vector machine (SVM) and a binary SVM. Experimental results on the Extended Cohn-Kanade dataset (CK+), indicate that the proposed algorithm reaches an average accuracy of 89.78% at a detection speed of less than 30 ms. The accuracy is comparable with state of-the-art emotion detection algorithms and outperforms these algorithms when detecting the emotions “Surprise”, “Fear”, “Disgust” and “Contempt” . The detection speed evaluation of the proposed algorithm was performed on a Windows 8.1 laptop with 8 GigaBytes of RAM and Intel Core i7-5500U CPU (2.40 GHz) [7]

In this paper, a novel method for facial expression recognition is proposed. The facial expression is extracted from human faces by an expression classifier that is learned from boosting Haar feature based Look-Up-Table type weak classifiers. The expression recognition system consists of three modules, namely - face detection, facial feature landmark extraction and facial expression recognition. The implemented system can automatically recognize seven expressions in real time that include “surprise”, “neutral”, “happiness”, “anger”, “sadness”, “fear” and “disgust”. [8]

There has been a great body of work with in-depth study in “Emotional recognition based on facial expressions”. In this paper, analysis and comparison of the state-of-the-art facial expression recognition techniques is done, alongwith proposals of some evaluation dimensions and discussion of possible directions for future research. [9]

Music plays a very important role in the daily lives of human beings and in advanced modern technologies. Generally, the user has to face the task of manually browsing through his/her playlist of songs to select. Here, an efficient and accurate model is proposed that would generate a playlist based on current emotional state and behaviour of the user. Existing methods for automating the playlist generation process are slow in computation, less accurate and sometimes even require use of additional hardware like sensors or EEG. Speech is the most ancient and natural way of expressing feelings, emotions and mood and its processing requires high computation, time, and cost. This proposed system based on real-time extraction of facial data as well as extracting audio features from songs to classify them into a specific emotion that will generate a playlist automatically such that the computation cost is relatively low. [10]

In this project we digitally sense the body temperature and the heart rate of a person using Arduino. Arduino is chosen because it is able to sense the environment by receiving input from a variety of sensors and can affect its surroundings by controlling motors, lights, and other actuators. The microcontroller on the board is programmed using the Arduino programming language. LM35 is used for sensing body heat/temperature. Body temperature is a basic parameter for monitoring and diagnosing human health. Heart beat sensor was used for sensing heart rate. This device will allow one to measure their mean arterial pressure (MAP) in about one minute and the accurate body temperature will be displayed on the Android. The system can be used to measure physiological parameters, such as Pulse rate and Heart rate (Systolic and Diastolic). [11]

This paper proposes a learning emotion recognition model, which consists of three stages: Feature extraction, subset feature and emotion classifier. Haar Cascade is used to detect the facial image, as the basis for the extraction of eyes and mouth regions, and then Sobel edge detection is used to obtain the characteristic value. Through Neural Network classifier training, six kinds of different emotional categories are obtained. Experiments were done using the JAFF database; which show that the proposed method has high classification performance. Experimental results show that the model proposed in this paper is consistent with the expressions from the learning situation of students in virtual learning environments. This paper demonstrates that emotion recognition based on facial expressions is feasible in distance education, permitting identification of a student's learning status in real time. Hence, it can help teachers to change teaching strategies in virtual learning environments according to the emotions of the students. [12]

Songs, as an expression medium, have always been a popular choice to depict and understand human emotions. Reliable emotion-based classification systems can go a long way in helping us parse their meaning. However, research in the field of emotion-based music classification has not yielded optimal results. In this paper, an effective cross-platform music player, EMP is proposed, which recommends music based on the user's real-time mood. EMP provides smart mood based music recommendation by incorporating the capabilities of emotion context reasoning within our adaptive music recommendation system. The music player contains three modules: Emotion Module, Music Classification Module and Recommendation Module. The Emotion Module takes the user's facial image as input and makes use of deep learning algorithms to identify their mood, which has an accuracy of 90.23 percent. The Music Classification Module makes use of audio features to achieve a remarkable result of 97.69 percent while classifying songs into 4 different mood classes. The Recommendation Module suggests songs to the user by mapping their emotions to the mood type of the song, taking into consideration the preferences of the user. [13]

Handwriting analysis is the technique used to understand a person in a better way through his/her handwriting. By examining the handwriting, we can develop a sketch which reflects the writer's emotional outlays, fears, honesty, mental state and many other personality traits. Emotions include the interpretation, perception and response of the feelings related to the experience of any particular situation. They are the ones which bridge thoughts, feelings and actions. The main objective of this paper is to analyze the handwriting characteristics like Baseline, Slant, Pen-Pressure, Size, Margin and Zone to determine the emotion levels of a person. This will help identifying those people who are emotionally disturbed or depressed and need psychological help to overcome such negative emotions. [14]

Emotion control is one of personality characteristics that can be detected through handwriting or graphology. One of the advantages is it may help the counselor that has difficulties in identifying the emotion of their counselee. This study is to explore the fuzzy technique for feature extraction in handwriting and then identify the emotion of the person. This study uses the baseline or slope of the handwriting in determining the level of emotion control whether it is very low, low, medium, high or very high, through Mamdani inference. [15]

Speech emotion recognition is a challenging task, and extensive reliance has been placed on models that use audio features in building well-performing classifiers. In this paper, we propose a novel deep dual recurrent encoder model that utilizes text data and audio signals simultaneously to obtain a better understanding of speech data. As emotional dialogue is composed of sound and spoken content, our model encodes the information from audio and text sequences using dual recurrent neural networks(RNNs) and then combines the information from these sources to predict the emotion class. This architecture analyzes speech data from the signal level to the language level, and it thus utilizes the information within the data more comprehensively than models that focus on audio features. Extensive experiments are conducted to investigate the efficacy and properties of the proposed model. Our proposed model outperforms previous state-of-the-art methods in assigning data to one of four emotion categories (i.e., angry, happy, sad and neutral) when the model is applied to the IEMOCAP dataset, as reflected by accuracies ranging from 68.8% to 71.8%. [16]

EEG signals measure the neuronal activities on different brain regions via electrodes. Many existing studies on EEG-based emotion recognition do not exploit the topological structure of EEG signals. In this paper, we propose a regularized graph neural network (RGNN) for EEG-based emotion recognition, which is biologically supported and captures both local and global inter-channel relations. Specifically, we model the inter-channel relations in EEG signals via an adjacency matrix in our graph neural network where the connection and sparseness of the adjacency matrix are supported by the neuroscience theories of human brain organization. In addition, we propose two regularizers, namely node-wise domain adversarial training (NodeDAT) and emotion-aware distribution learning

(EmotionDL), to improve the robustness of our model against cross-subject EEG variations and noisy labels, respectively. To thoroughly evaluate our model, we conduct extensive experiments in both subject-dependent and subject-independent classification settings on two public datasets: SEED and SEED-IV. Our model obtains better performance than competitive baselines such as SVM, DBN, DGCNN, BiDANN, and the state-of-the-art BiHDM in most experimental settings . Our model analysis demonstrates that the proposed biologically supported adjacency matrix and two regularizers contribute consistent and significant gain to the performance. Investigations on the neuronal activities reveal that pre-frontal, parietal and occipital regions may be the most informative regions for emotion recognition, which is consistent with relevant prior studies. In addition, experimental results suggest that global inter-channel relations between the left and right hemispheres are important for emotion recognition and local inter-channel relations between (FP1, AF3), (F6, F8) and (FP2, AF4) may also provide useful information. [17]

Emotion recognition is a rapidly growing research domain in recent years. Unlike humans, machines lack the abilities to perceive and show emotions. But human-computer interaction can be improved by automated emotions recognition, thereby reducing the need of human intervention. In this paper, four basic emotions (Anger, Happy, Fear and Neutral) are analyzed from emotional speech signals. Signal processing methods are used for obtaining the production features from these signals. Source features the instantaneous fundamental frequency (F0), system features the formants and dominant frequencies, zero-crossing rate (ZCR), and the combined features signal energy are used for the analyses. F0 is obtained using zero-frequency filtering (ZFF), and formants and dominant frequencies using LP spectrum. Short-time signal energy (STE) and ZCR are obtained in the voiced and unvoiced regions using a rectangular window of 200 samples. Two databases, German and Telugu Emotion Databases are used to cross-validate the results. Distinct differences are observed between high-arousal emotions (Anger and Happy) and Neutral emotion. Results indicate overlap between Anger and Happy emotions. But distinct differences are observed in the features for Happy/Anger and Fear, and between Happy and Anger emotions which is otherwise a challenging problem. The insights gained may be helpful in a range of applications. [18]

In this paper, the aim was to reveal temporal variations in videos that were deemed impossible to see with the naked eye and display them. The method that is proposed, which is called Eulerian Video Magnification, takes a standard video sequence as input, and applies spatial decomposition. This is then followed by temporal filtering to the frames. The resulting signal is then amplified to reveal hidden information that is invisible to the naked eye. Using this method, visualization of the flow of blood as it fills the face is done and it is also amplified to reveal small motions. This proposed method can run in real time to show phenomena occurring at temporal frequencies set by the user [19]

2.2 Suggestions based on literature review

On the basis of the above literature review, we have inferred that for the Emotion detection phase four different approaches can be used, namely -

- **By using Handwriting Recognition** - Handwriting analysis is the technique used to understand a person in a better way through his/her handwriting. Handwriting characteristics such as Baseline, Margin, Pen-Pressure, Slant, Zone, and Size can help to determine the emotion of a person. Out of all the handwriting characteristics, the Baseline characteristic is commonly used to understand the emotion of a person.

  The line on which most of the letters rest is known as baseline. There are four most common baselines namely Ascending, Descending, Straight and Wavy Baseline whose significant emotional characteristics are given below : -
  - ➢ *Descending Baseline* - Depression, pessimistic, tired.
  - ➢ *Ascending Baseline* - Optimistic, tensed and over-disciplined
  - ➢ *Straight Baseline* - Healthy balance between optimistic and pessimistic, emotional restraint
  - ➢ *Wavy Baseline* - Emotional roller coaster, out of bounds, emotionally unsteady.
- **By Speech Processing** - Speech processing can be used to detect the emotion of a user by converting the speech signal into a graph and then analysis is done based on the graph. One of the main feature attributes considered was

the peak-to-peak distance obtained from the graphical representation of the speech signals. As emotional dialogue is composed of sound and spoken content, there is a given model that encodes the information from audio and text sequences using dual recurrent neural networks(RNNs) and then combines the information from these sources to predict the emotion class. This architecture analyzes speech data from the signal level to the language level, and it thus utilizes the information within the data more comprehensively than models that focus on audio features.

- **By estimating the Heart Rate** - Heart Rate can be estimated by imperceptible motions of the head caused by blood flow or by observing the blood oxygenated haemoglobin which absorbs the green light and measures the heart rate by detecting the variation of the green color intensity of the person's face.

  However by estimating the heart rate and ranging it in order to find the current emotion of the user is somewhat difficult. As characteristics of heart rate vary from person to person as it depends on age, gender, fitness level, overall lifestyle etc.

- **By using Facial Input** - Emotions can be detected by extracting features from the facial input. As when people are by themselves they tend to express their true emotions. For the purpose of extracting features various approaches can be used. However from the above review we draw the inference that before extracting the feature localization of the facial input can be done and this localized face region can then be forwarded for further processing. For localizing the facial region Haar cascade method suits the best as there will be a small number of classes for classification and for facial feature extraction these regions can be fed forwarded to a CNN model.

Since we want our system to be automatic and also depending upon the mood of the user, the user may not be willing to speak or write to our system. Hence, we have excluded Speech processing and Handwriting Processing.

As characteristics of heart rate may vary from person to person as it depends on age, gender, fitness level, overall lifestyle etc, it is difficult to classify a particular heart beat range into a particular emotion, we have excluded the Heart rate Estimation for Emotion Recognition.

As a result of all the above inconsistencies in the above mentioned methods, Recognition of Emotion from Facial Input remains the best method for Emotion Recognition.

2.3 Current Scenario

As there is no existing system that can help uplift the emotion of a person and in turn help in keeping better mental health. The closest match to a "system" that has been laid by the Indian Govt. as a foundation for mental health is as below :-

ROADMAP FOR MENTAL HEALTH AWARENESS

For the large Indian population to be involved in its own mental health, the only way forward is through enhancing mental health awareness which will generate its own demand. With rising awareness, it can be expected that early recognition and access to treatment will follow, as will the adoption of preventive measures. It can also be expected that with enlarging awareness in a democratic society, advocacy, leveraging of political will, funding, and cross-synergies shall follow. It is envisaged that bulk of the awareness contributions shall flow from the following six platforms

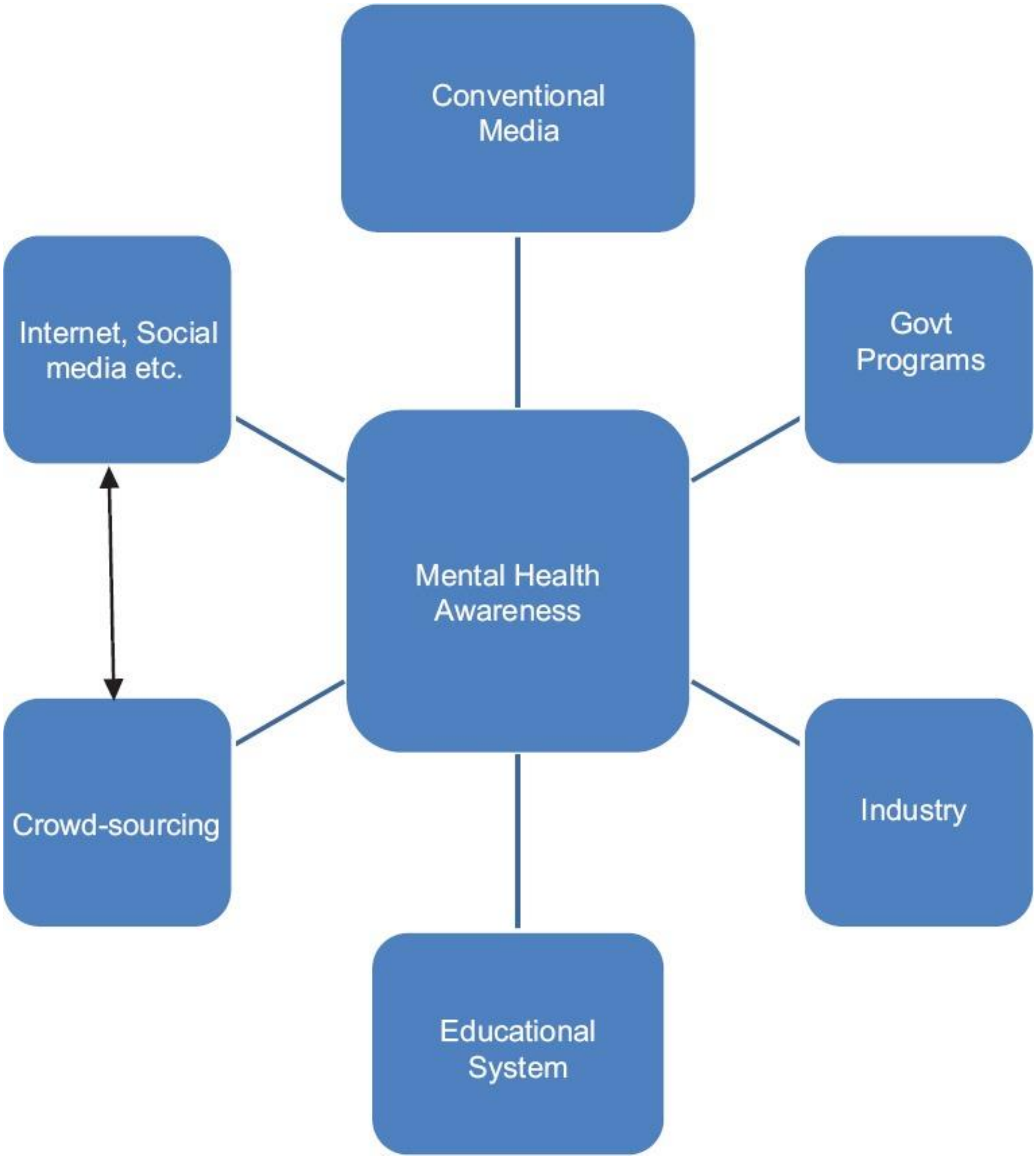


Fig 2.1: Roadmap for Mental Health Awareness

Source : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479084/

Conventional media

Media has been the cornerstone of the previous action in the field of mental health awareness. Celebrity endorsements, like the recent one by actress Deepika Padukone who shared her experience of depression, together with succinct tag-lines of advertisements and content-rich narrations and documentaries; have been the mainstay of media drives so far.

Making evidence-based mental health information easily available to journalists and other content providers like internet portals from trusted and reliable sources like Indian Psychiatry Society, research organizations, medical colleges, etc., through their websites is a relatively simple step. Accessibility of simply translated jargon-free content in various regional languages in written and spoken forms will go a long way. It also behoves professionals in the mental health domain to take the lead in engaging and partnering with the media. Encouraging recovered patients to make their success stories accessible to all shall make good the paucity of authentic narratives.

Government Programs

Despite some caviling about the quantum, the government remains the biggest single spender in the mental health sector. While most new interventions remain isolated and confined to urban areas, it is only the public health system through large programs which can reach the rural masses. Apart from the National and District Mental Health Programs, the National Rural Health Mission is on its way to becoming the vehicle for delivering mental health as a part of integrated primary care at the cutting edge of the public healthcare system. Seeing that it partners with existing private and alternative care providers in a nonthreatening manner, shall help such large interventions synergize and succeed.

Education System

Most chronic and debilitating mental illnesses have their onset before 24 years of age when most are a part of the educational system. From including mental health narratives in curricula toward, de-stigmatization, removing discrimination and early detection, to empowering stakeholders for early detection and simple interventions; the educational system yields myriad opportunities for enhancing mental health awareness.

Industry

The organized sector suffers significant loss of effective workforce through mental ill-health. Not only as a part of corporate social responsibility but also to maintain productivity, it becomes important to engage with mental health awareness in a concerted fashion.

Internet, social media and cell phones

Hand-held devices and the social media can truly be game-changers in the propagation of effective mental health interventions through focussed amplification, and not just in increasing information. With the greater utilization of big data, the understanding of subtle and distributed patterns over large volumes shall inform decision making.

Crowd-sourcing

The ultimate convergence of information and technology in a free society results in crowd-sourcing which breaks down barriers of geography, historical inequities, and economies of scale. It is the true involvement of communities real and virtual, harnessed to make a change. Thus, dynamic ideas of individuals can synergize with the success stories of nongovernmental organizations to amplify them across geographies and time. Crowd-funding is a successful model in testing radical ideas which flounder outside the mainstream. [1]

How our model can help reduce the risk of having negative emotions

Since mental health is a big issue and keeping oneself mentally healthy is important, reducing negative emotions is an integral part of this. So in order to achieve this our proposed model can help and work alongside the above mentioned existing methods of dealing with mental health issues which in turn boils down to reducing negative emotions.

The system proposed can recognize a user’s emotion through the variety of characteristic features and signs that a person may show while experiencing said emotion.

The patterns that the system uses to recognize an emotion from a face can be given as follows:

Happiness

A Genuine, fully expressed happy face includes:

- Wide smile with open mouth - you should see teeth.
- "Crow's feet" around the eyes (wrinkles around the eyes)
- Raised cheeks
- The eyes squint some
- Possible wrinkles around the nose.

Sadness

The simple way to remember the features of a sad face is to think that it feels like the face "melts":

- A frown - we slant and raise our inner eyebrows. (a hard thing to fake)
- The lip's corners are pulled down
- There might be a tension in the neck and chin area (holding back the tears).

When we fake sadness we tend to overdo it (like kids) - we stick out the lower lip and wear a "sad smile". Real sadness however is hard to fake:

1. Real frowns require fine control of the brows, we need to slant and raise only our inner brows.
2. It's often a "quite" expression, it's as if someone is "switched off" - he might gaze down and look somewhat contemplative.

Anger

The keyword here is tension - tension in the eyebrows, jaws, lips and around the eyes.

- Closed 'V shaped' eyebrows.
- Open and square mouth Or closed mouth with tense chin and jaws.
- Squinty eyes with fixed icy stare.
- "Flaring" nose in overdramatic individuals (kids...)

Our brows get closer and in downward V shape when we're truly angry. There is a lot of tension in the central point between the brows and a fixation with the eyes on the target of the rage.

Even if you see a smile, but also observe a tense forehead and a fixed gaze - you can be sure something is wrong. [20]

Further after recognizing an emotion, the system now addresses the “need” of that specific emotion. For example: Suggesting calming exercises to the angry one, Suggesting music to the sad one. Further establishing that our system can change its output depending on what the user’s input is. So the flexibility of our system helps improve the emotional and mental state of a user. Our system can hence be used as a tool that can help users improve their mental being.

Thereby the following sets of activities can be claimed to be better suited to enhance the negative as well as positive emotions.

- ***Cool Off with Exercise*** - A great outlet to reduce tension is physical activity: using one’s anger as fuel for a healthier lifestyle.
- ***Listen to Music*** - The meter, timber, rhythm and pitch of music are managed in areas of the brain that deal with emotions and moods.
- ***Listen to Audiobooks*** - Audio books can captivate the imagination, allowing listeners to create a whole world at once within and outside themselves.
- ***Perform a quick calming breathing activity*** - Breathing is the number one and most effective technique for reducing anger and anxiety quickly.
- ***Watch funny videos*** - When it comes to relieving stress, more giggles and guffaws are just what the doctor orders as laughter enhances one's intake of oxygen-rich air, stimulates the heart, lungs and muscles, and increases the endorphins that are released by the brain.
- ***Watch a movie*** - Cinema therapy allows one to use the effect of imagery, plot, music, etc. in films on their psyche for insight, inspiration, emotional release or relief and natural change.
- ***Meditation*** - Meditation as a therapy can help positive physical and psychological changes. In some people, meditation can help reduce asthma, allergies, high blood pressure, and pain. Many advocates of meditation point out that the state of the mind affects the state of the body, and that a peaceful mind can enable the body to heal itself.

Chapter - 3 : Design & Implementation

3.1 Overview of the Proposed System.

The proposed system is an Emotion detection and improving Assistant System that is just the right solution for the problems and shortcomings of the problems that include issues like loneliness, unwillingness to participate in group activities and stepping out of one’s comfort zone to tackle their problems. This is done by using a facial recognition system that can detect the emotion of the user and based on that emotion several activities are going to be outputted to the user. This is greatly beneficial as there are several reports that people do not like to engage in group activities or social gatherings, especially in India where it is misunderstood or seen as a taboo to have mental problems. Since this system is targeted towards people who do not like to engage in group activities.

Considering the application of this system, one might think that the system is only capable of suggesting activities for the negative emotions. While the system is primarily designed to do this, that is not mainly the case as there are a list of several activities that can be beneficial to people who are currently not having any negative emotions as well. So this is further helpful for keeping positive emotions in check so that they do not change into negative ones anytime soon.

3.2 Methodology

3.2.1 Functional requirements

The following are the system functional requirements of the system;

- The system should be able to identify, recognize and authenticate the face of the person.
- The system should be able to detect and recognize the emotion accurately.
- The system should be able to call a group of activities and curate them for the user in text and in voice.

- The system should be able to output a list of activities based on the emotion being shown by the user.

- The system should be accessible by a blind person as easily as possible.

3.2.2 Non-functional requirements

Non-functional requirements define the overall qualities or attributes of the resulting system. Non-functional requirements are constraints of the product developed that must meet to make the system useful. These includes the following:

- **Performance** – The system uses short response time, high throughput and easy Utilization.

- **Usability** - The system is easy to use for any user.

- **Scalability**- The system can be extended to increase total throughput under an increased load when resources (typically hardware) are added.

- **Availability**- the system is available for service when requested by the user at any time.

- **Reliability**- the system performs its required functions under stated conditions for a specific period of time.

- **Maintainability**- the system is maintainable to incorporate other functionalities.

- **Security**- Unauthorized access to the system is not allowed.

3.2.3 Hardware Requirements.

The list below is the hardware components required to complete the project design and implementation of the Emotion based assistant service system:-

- A computer capable of running TensorFlow-GPU (NVidia GPU only) and python
- Webcam with good lighting conditions and clear image.

- Proper Microphone for clear audio input from the user.

3.2.4 Phases of implementation

The procedural development of proposed system components are grouped into 4 different phases, based on the function carried out by the modules of the prototype as briefly described in this section.

1. Facial Image Region localization
2. Facial feature extraction
3. Classification of emotion
4. Activity Suggestion
    a) For Happy
    b) For Sad
    c) For Angry
    d) For Neutral

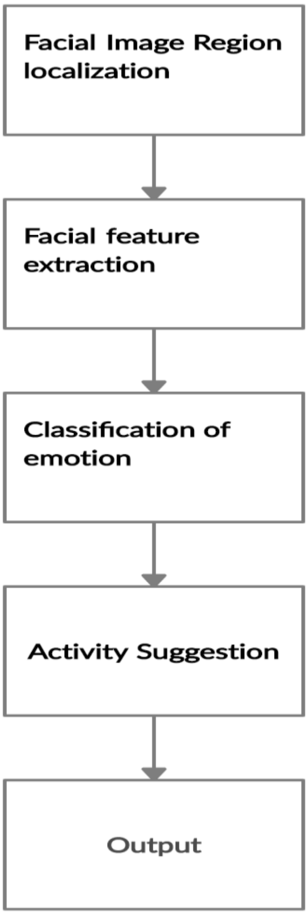

Facial Image Region localization
Facial feature extraction
Classification of emotion
Activity Suggestion
Output

Figure 3.1: Block diagram of the Phases of Implementation

3.2.4.1 Facial Image Region Localization

Firstly, image acquisition is done by the USB Image capture camera which captures the facial input in real-time when it is activated. The positioning of the camera will be structured to get the image of the face that will be further used to localize the facial region which in turn can be forwarded for feature extraction.

After the input is captured, before localization; preprocessing of the facial input is done in order to decrease the computation time and is sent to the system. The main purpose of the pre-processing is to increase the efficiency of Facial Feature Extraction phase which includes the set algorithms applied on the video input to enhance the accuracy while obtaining the correct face value as required. It is an important phase in the system

The pre-processing techniques used in the Facial Image Region localization are as follows:

a) ***Gray scale conversion***: Grayscale images consist of only gray tones of colors, which are only 256 steps, which means that there are only 256 gray colors (0 - 255). The main characteristic of grayscale images is the equality of the red, green, and blue color levels. The color code will be like RGB(R,R,R), RGB(G,G,G), or RGB(B,B,B) where 'R,G,B' is a number between 0 and 255 individually. Gray Scale conversion is used to reduce the computation power needed to process the input video(frames). This is done by the following function.

```
gray = cv2.cvtColor(frame,cv2.COLOR_BGR2GRAY)
```

b) ***Localization of Facial Region*** : For the detection of the facial region, we have used Haar Cascade since we only require two classes i.e Face or No face rather than using other methods such as YOLO, CNN which uses high computational power.

Haar Cascade method

Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

Here, we will work with face detection. Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract the features from it. For this, Haar features shown in the below image are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting the sum of pixels under the white rectangle from the sum of pixels under the black rectangle.

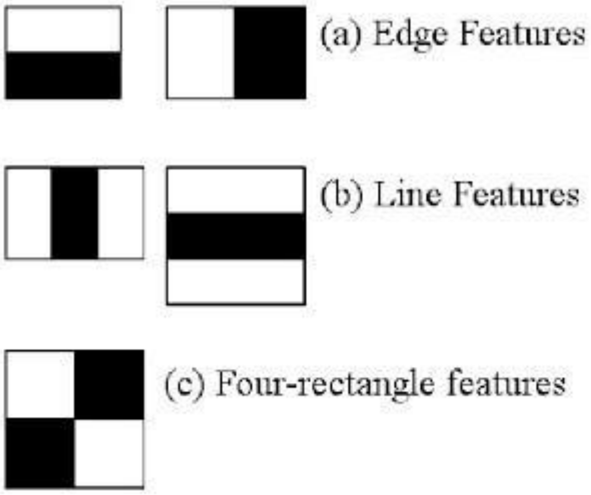


Figure 3.2: Haar feature types.

Source: Rapid Object Detection using a Boosted Cascade of Simple Features Paul Viola Michael Jones viola@merl.com mjones@crl.dec.com Mitsubishi Electric Research Labs Compaq CRL 201 Broadway, 8th FL One Cambridge Center Cambridge, MA 02139 Cambridge, MA 02142

Now, all possible sizes and locations of each kernel are used to calculate lots of features. For each feature calculation, we need to find the sum of the pixels under the black and white rectangles. To solve this, they introduced the integral image. However large the image is, it reduces the calculations for a given pixel to an operation involving just four pixels. It makes things ultra-fast.

But among all these features we calculated, most of them were irrelevant. For example, consider the image below. The top row shows two good features. The first feature selected seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose. But the same windows applied to cheeks or any other place is irrelevant. So, we need to select the best features out of 160000+ features. This is achieved by **Adaboost**.

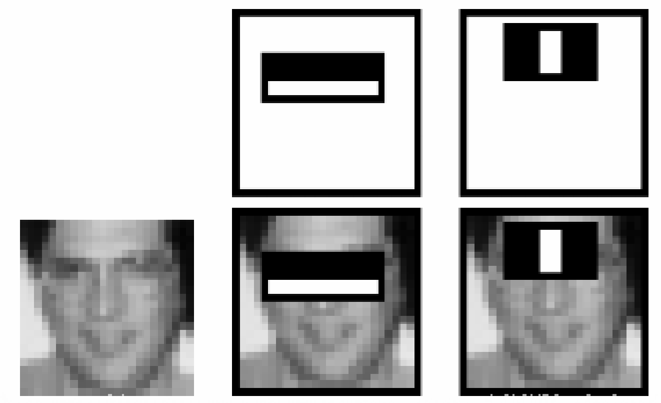


Figure 3.3: Haar Features being detected

Source: Rapid Object Detection using a Boosted Cascade of Simple Features Paul Viola Michael Jones viola@merl.com mjones@crl.dec.com Mitsubishi Electric Research Labs Compaq CRL 201 Broadway, 8th FL One Cambridge Center Cambridge, MA 02139 Cambridge, MA 02142

For this, we apply each and every feature on all the training images. For each feature, it finds the best threshold which will classify the faces to positive and negative. Naturally, there will be errors or misclassifications. We select the features with minimum error rate, which means they are the features that most accurately classify the face and non-face images.
(*Note*: The process is not as simple as this. Each image is given an equal weight in the beginning. After each classification, weights of misclassified

images are increased. Then the same process is done. New error rates, as well as new weights are calculated. The process is continued until the required accuracy or error rate is achieved or the required number of features are found).

The final classifier is a weighted sum of these weak classifiers. It is called weak because it alone cannot classify the image, but together with others forms a strong classifier. The paper says even 200 features provide detection with 95% accuracy. Their final setup had around 6000 features.
(Imagine a reduction from 160000+ features to 6000 features. That is a big gain).

After an image is taken, 24 * 24 windows are chosen and we apply 6000 features to it and check if it is face or not. This is a little inefficient and time consuming.

In an image, most of the image is non-face region. So it is a better idea to have a simple method to check if a window is not a face region. If it is not, discard it in a single shot, and don't process it again. Instead, focus on regions where there can be a face. This way, we spend more time checking possible face regions.

For this, they introduced the concept of”**Cascade of Classifiers”**. Instead of applying all 6000 features on a window, the features are grouped into different stages of classifiers and applied one-by-one. (Normally the first few stages will contain very many fewer features). If a window fails the first stage, discard it. We don't consider the remaining features on it. If it passes, apply the second stage of features and continue the process. The window which passes all stages is a face region.

The authors' detector had 6000+ features with 38 stages with 1, 10, 25, 25 and 50 features in the first five stages. (The two features in the above image are actually obtained as the best two features from Adaboost). According to the authors, on average 10 features out of 6000+ are evaluated per sub-window.

Haar-like cascade classifiers were first used in face detection. A number of Haar-like features in a default window are extracted. In our experiment, 15 features were used. The exhaustive set of Haar-like features in a default window is over-complete and much more than the number of pixels. AdaBoost is an effective classifier to train from a great number of features. It selects a small number of efficient features to build a weak classifier, stump classifier or classification and regression tree classifier. We can see that the weak classifier largely depends on the selection of positive and negative samples. So how to choose samples is a core problem in AdaBoost's training. Weak classifiers are then combined to form a strong classifier. The structure of a strong classifier is shown in the following figure.

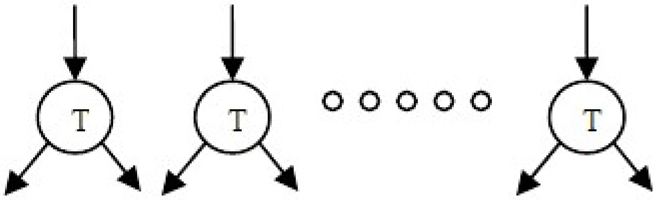


Figure 3.4: the structure of AdaBoost

Source: Rapid Object Detection using a Boosted Cascade of Simple Features Paul Viola Michael Jones viola@merl.com mjones@crl.dec.com Mitsubishi Electric Research Labs Compaq CRL 201 Broadway, 8th FL One Cambridge Center Cambridge, MA 02139 Cambridge, MA 02142

To reduce computation time in the detection stage, a cascade structure is adopted. The total time is reduced immensely by processing large areas of the input image via simple classifiers and processing the rest of the input image via complex classifiers.

In our system, we have trained two haar cascade classifiers. We have used one classifier in the Facial Region Localization phase and the other one for the same purpose but for a different phase which is explained in the later section.

c) **Rectangle:** A rectangle is drawn over the face region that is detected using Haar Cascade.

d) **ROI (Region of Interest):** The rectangular region is then cut out of the frame

e) **Resize**: To decrease the computation time, the input video frame is resized to a particular size.

3.2.4.2 Facial feature extraction

In order to build the emotion classifier, the Facial Feature Extraction from the detected face is done by Convolution Neural Network (CNN).

Convolution Neural Network (CNN)

In machine learning, a convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation. Convolutional networks were inspired by biological processes and are variations of multilayer perceptrons designed to use minimal amounts of preprocessing. They have wide applications in image and video recognition, recommender systems and natural language processing. The convolutional neural network is also known as shift invariant or space invariant artificial neural network (SIANN), which is named based on its shared weights architecture and translation invariance characteristics. [22]

VGG16 is a simpler convolutional neural network model, since it does not use hyper parameters. It was one of the famous models submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3*3 kernel-sized filters one after another.

VGG 16 Architecture - The input to the cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.

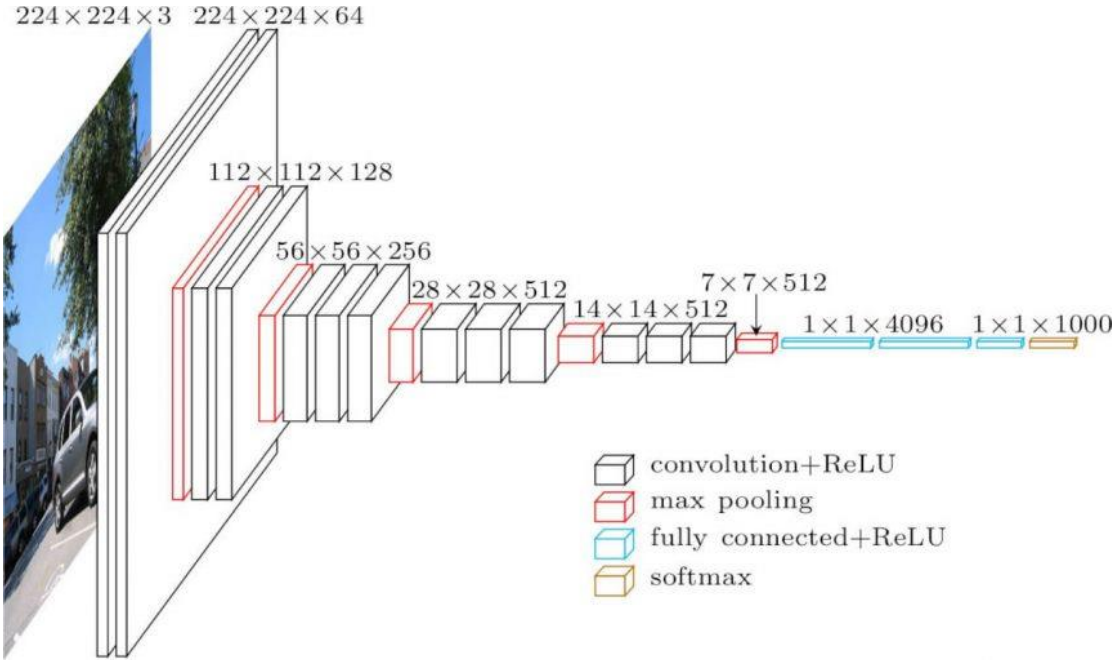

$224 \times 224 \times 3$
$224 \times 224 \times 64$
$112 \times 112 \times 128$
$56 \times 56 \times 256$
$28 \times 28 \times 512$
$14 \times 14 \times 512$
$7 \times 7 \times 512$
$1 \times 1 \times 4096$
$1 \times 1 \times 1000$
convolution+ReLU
max pooling
fully connected+ReLU
softmax

Figure: 3.5 VGG Architecture

Source: https://neurohive.io/en/popular-networks/vgg16/

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

All hidden layers are equipped with the rectification (ReLU) non-linearity. It is also noted that none of the networks (except for one) contain Local Response Normalisation (LRN), such normalization does not improve the performance on the ILSVRC dataset, but leads to increased memory consumption and computation time. [23]

There are four main operations in the Convolution Neural Network shown in the figure above:

1. **Convolution**: The primary purpose of Convolution in the case of a CNN is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data. The convolution layer's parameters consist of a set of learnable filters. Every filter is small spatially (along width and height), but extends through the full depth of the input volume. For example, a typical filter on a first layer of a CNN might have size 3*5*5 (i.e. images have depth 3 i.e. the color channels, 5 pixels width and height). During the forward pass, each filter is convolved across the width and height of the input volume and computes dot products between the entries of the filter and the input at any position. As the filter convolves over the width and height of the input volume it produces a 2-dimensional activation map that gives the responses of that filter at every spatial position. Intuitively, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some color on the first layer, or eventually entire honeycomb or wheel-like patterns on higher layers of the network. Now, there will be an entire set of filters in each convolution layer (e.g. 20 filters), and each of them will produce a separate 2-dimensional activation map.

   A filter convolves with the input image to produce a feature map. The convolution of another filter over the same image gives a different feature map. Convolution operation captures the local dependencies in the original image. A Convolution Neural Network learns the values of these filters on its own during the training process (although parameters such as number of filters, filter size, architecture of the network etc. are still needed to specify before the training process). The greater the number of filters, the more image features get extracted and the better the CNN becomes at recognizing patterns in unseen images. The size of the Feature Map (Convolved Feature) is controlled by three parameters:

   a) ***Depth***: Depth corresponds to the number of filters we use for the convolution operation.

b) ***Stride***: Stride is the size of the filter, if the size of the filter is 5*5 then stride is 5.

c) ***Zero-padding***: Sometimes, it is convenient to pad the input matrix with zeros around the border, so that filter can be applied to bordering elements of the input image matrix. Using zero padding size of the feature map can be controlled. [22]

2. **Exponential Linear Unit (ELU)**: This activation function fixes some of the problems with ReLUs and keeps some of the positive things. For this activation function, an alpha α value is picked; a common value is between 0.1 and 0.3. The equation is given below:

$$\mathrm{ELU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x < 0 \end{cases} \quad \ldots(3.1)$$

If we input an x-value that is greater than zero, then it is the same as the ReLU i.e the result will be a y-value equal to the x-value. But this time, if the input value, x is less than 0, we get a value slightly below zero. The y-value we get depends both on our x-value input, but also on a parameter alpha α, which we can adjust as needed. Furthermore, we introduce an exponential operation (e^x , which means the ELU is more computationally expensive than the ReLU. The ELU function is plotted below with an α value of 0.2.

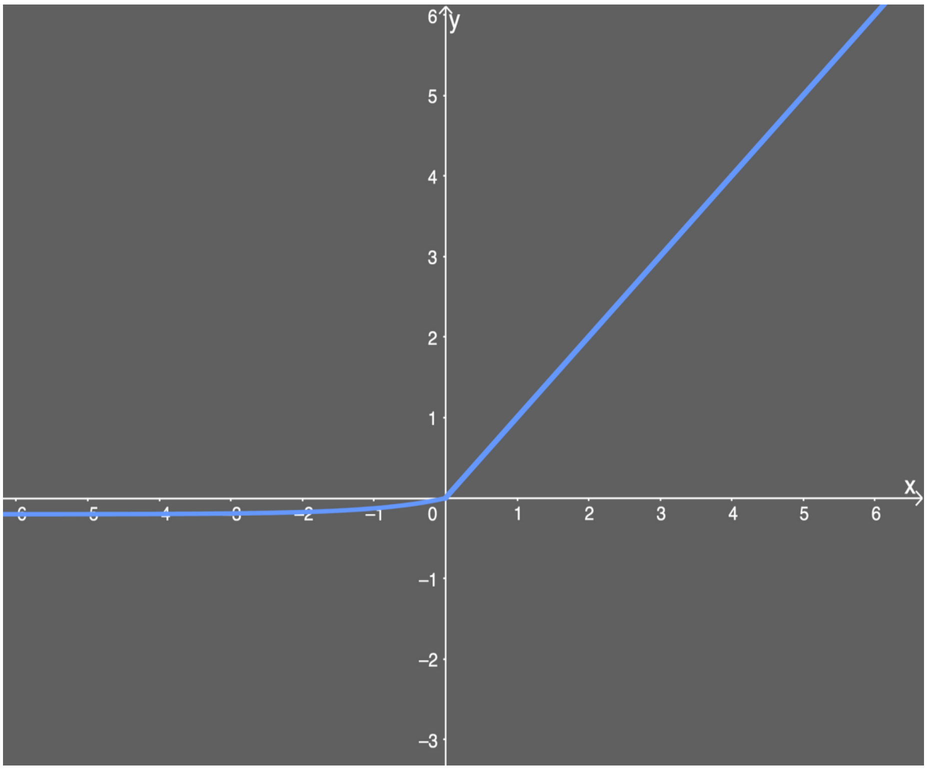


Figure 3.6: The plot for the ELU activation function

Source: https://mlfromscratch.com/activation-functions-explained/#/

Derivative of the ELU:-

$$\text{ELU`}(x) = \begin{cases} 1 & \text{if } x > 0 \\ \text{ELU}(x) + \alpha & \text{if } x < 0 \end{cases} \qquad \text{….(3.2)}$$

The y-value output is 1 if x is greater than 0. The output is the ELU function (not differentiated) plus the alpha value, if the input x is less than zero.

The plot for it looks like this:

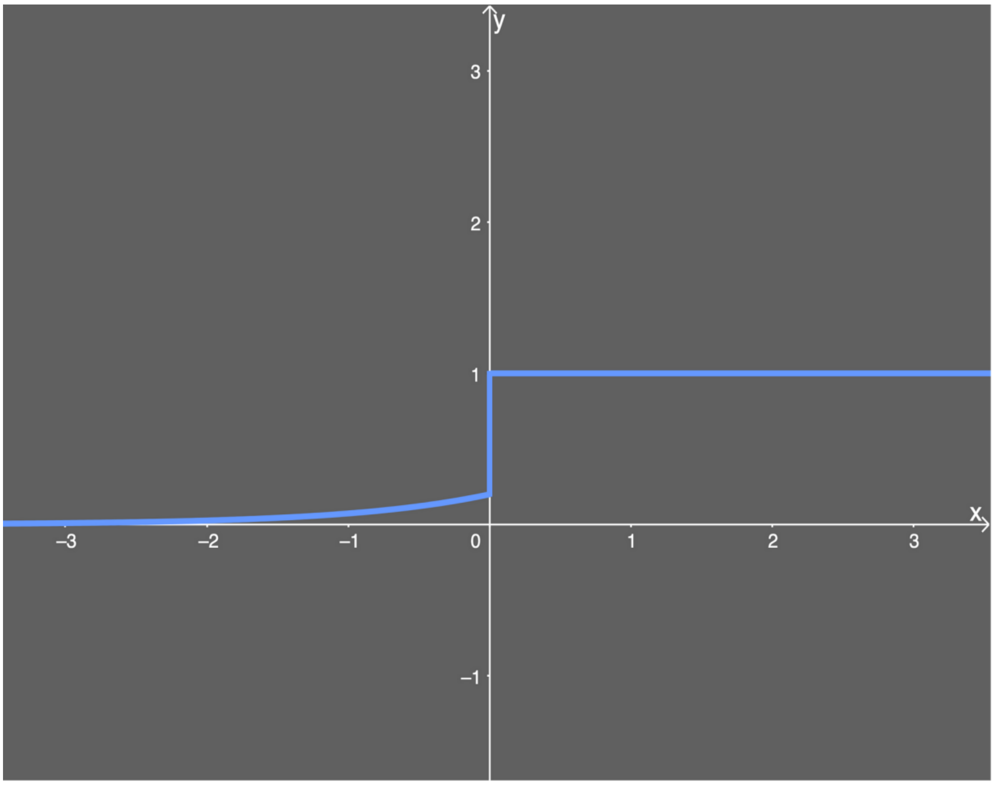


Figure 3.7: The ELU activation function differentiated

Source: https://mlfromscratch.com/activation-functions-explained/#/

We avoid the dead ReLU problem here, while still keeping some of the computational speed gained by the ReLU activation function – that is, we will still have some dead components in the network.

Pros

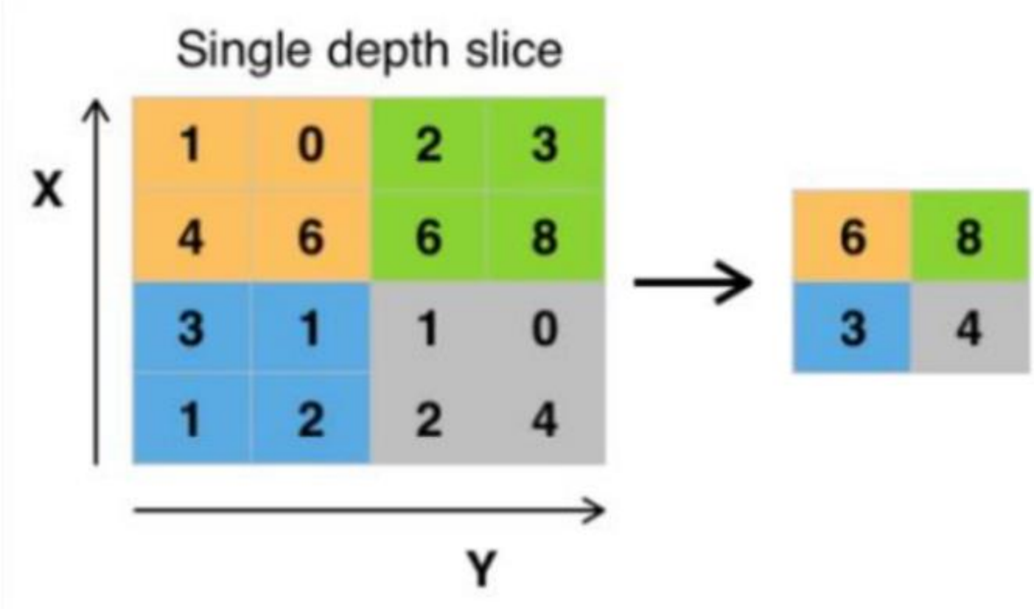
- Avoids the dead relu problem.
- Produces negative outputs, which helps the network nudge weights and biases in the right directions.
- Produce activations instead of letting them be zero, when calculating the gradient.

Cons

- Introduces longer computation time, because of the exponential operation included
- Does not avoid the exploding gradient problem
- The neural network does not learn the alpha value

3. **Pooling** (sub-sampling) Spatial Pooling (also called sub sampling or down sampling) reduces the dimensionality of each feature map but retains the most important information. Spatial Pooling can be of different types: Max, Average, Sum etc. In the case of Max Pooling, a spatial neighborhood (for example, a 2×2 window) is defined and the largest element is taken from the rectified feature map within that window. In case of average pooling the average or sum of all elements in that window is taken. In practice, Max Pooling has been shown to work better. Max Pooling reduces the input by applying the maximum function over the input xi. Let m be the size of the filter, then the output calculated as follows:

$$M(xi) = max \{ xi +k, l[k] <= m/2, [l] <=m/2 \; k, l \in N\}$$



1	0	2	3
4	6	6	8
3	1	1	0
1	2	2	4

6	8
3	4

Figure 3.8: Max Pooling

Source: Facial Expression Recognition with Convolutional Neural Networks Arushi Raghuvanshi Stanford University arushir@cs.stanford.edu Vivek Choksi Stanford University vchoksi@cs.stanford.edu

The function of Pooling is to progressively reduce the spatial size of the input representation. In particular, pooling Makes the input representations (feature dimension) smaller and more manageable

- Reduces the number of parameters and computations in the network, therefore, controlling over-fitting
- Makes the network invariant to small transformations, distortions and translations in the input image (a small distortion in input will not change the output of Pooling.
- Helps us arrive at an almost scale invariant representation. This is very powerful since objects can be detected in an image no matter where they are located. [22]

4. **Classification (Multilayer Perceptron):** The Fully Connected layer is a traditional Multi-Layer Perceptron that uses a softmax activation function in the output layer. The term "Fully Connected" implies that every neuron in the previous layer is connected to every neuron on the next layer. The output from the convolutional and pooling layers represent high-level features of the input image. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes based on the training dataset. Softmax is used for activation functions. It treats the outputs as scores for each class. In the Softmax, the function mapping stayed unchanged and these scores are interpreted as the un-normalized log probabilities for each class. Softmax is calculated as:

$$f(\mathrm{z})_{\mathrm{j}} = \frac{exp\,(\mathrm{zj})}{\sum_{K=1}^{K} exp\,(zk)} \quad .... (3.3)$$

Where j is index for image and K is number of total facial expression class. Apart from classification, adding a fully-connected layer is also a (usually) cheap way of learning non-linear combinations of these features. Most of the features from convolutional and pooling layers may be good for the classification task, but combinations of those features might be even better. The sum of output probabilities from the Fully Connected Layer is 1. This is ensured by using the as the activation function in the output layer of the Fully Connected Layer. The Softmax function takes a vector of arbitrary real-valued scores and squashes it to a vector of values between zero and one that sums to one. [22]

3.2.4.2.1 Preprocessing and Training

1. **Preprocessing of dataset -** For the training of our classifier we have used the FER 2013 dataset. The Facial Expression Recognition 2013 (FER-2013) Dataset classify facial expressions from 35,685 examples of 48*48 pixel grayscale images of faces. Images are categorized based on the emotion shown in the facial expressions (happiness, neutral, sadness, anger, surprise, disgust, fear). As our classifier will be classifying four different types of emotions which are happiness, neutral, sadness, anger. We need to further process the FER2013 dataset to fit our needs. To do this we first trim the dataset i.e - removing surprise, disgust, fear data out of it. After removing, our dataset now contains 21055 items which will be further augmented in the next phase.
2. **Data augmentation** - Data augmentation encompasses a wide range of techniques used to generate “new” training samples from the original ones by applying random jitters and perturbations (but at the same time ensuring that the class labels of the data are not changed). This is done to expand our existing dataset.

   Our goal when applying data augmentation is to increase the generalizability of the model.

   For example, we can obtain augmented data from the original images by applying simple geometric transforms, such as random:

   - Translations
   - Rotations
   - Changes in scale

- Shearing
- Horizontal (and in some cases, vertical) flips

Data augmentation or generation is done both for training as well as for validation.

```
    Function DATA_GENERATOR(){
        1.Take an input image
        2. Apply Rescale
        3. Set Rotation range to 30
        4. Set Shear range to 30
        5. Set Zoom range to 30
        6. Set Width shift range to 30
        7. Set Height shift range to 30
        8. Activate Horizontal flip
        9. Set Fill Mode to nearest
Return the DATA_GENERATOR architecture
}
Function TRAIN_GENERATOR = ()
{
    1. Take DATA_GENERATOR architecture
    2. Set train_data directory
    3. Set color_mode to 'grayscale',
    4. Set target_size to (img_rows,img_cols),
    5. Set batch_size to batch_size,
    6. Set class_mode to 'categorical',
    7. Apply shuffle to True
Return the TRAIN_GENERATOR architecture
}
```

3. **Model Selection (Sequential):** A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor.

```
Function Layer(number of filters){
    1. Set Model to Sequential
    2. Create a convolution layer using conv2D with kernel size 3
    3. Apply elu activation function to it
    4. Apply Batch Normalization
    5. Repeat step 2 till step 4
    6. Apply Pooling using MaxPooling2D
    7. Activate Dropout
    8. Repeats step 1 till step 7 but double the batch size
    9. Repeat step 8 twice
    10. Add Flatten
    11. Apply softmax activation function to it
Return the Layer architecture
}
```

Activation: Applies an activation function to an output.

Dense: Dense layer is the regular deeply connected neural network layer. It is the most common and frequently used layer. Dense layer does the below operation on the input and returns the output.

output = activation(dot(input, kernel) + bias)

where, *input* represents the input data

kernel represents the weight data

dot represent numpy dot product of all input and its corresponding weights

bias represent a biased value used in machine learning to optimize the model

activation represents the activation function.

Flatten: Flatten is used to flatten the input. For example, if flatten is applied to layer having input shape as (batch_size, 2,2), then the output shape of the layer will be (batch_size, 4)

Flatten has one argument as follows

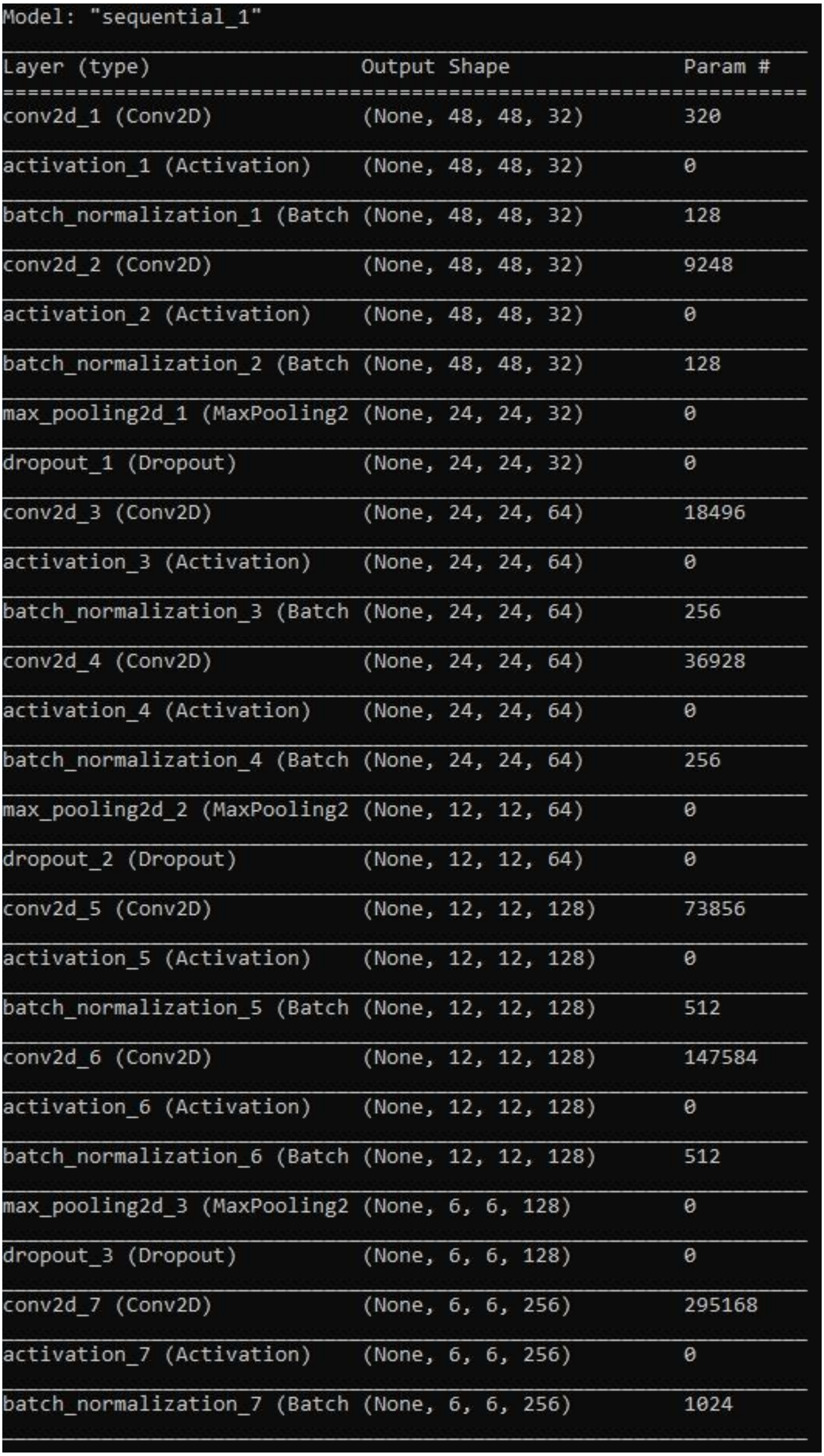
keras.layers.Flatten(data_format = None)

Conv2D: Conv2D is a 2D Convolution Layer, this layer creates a convolution kernel that is wind with layers input which helps produce a tensor of outputs.

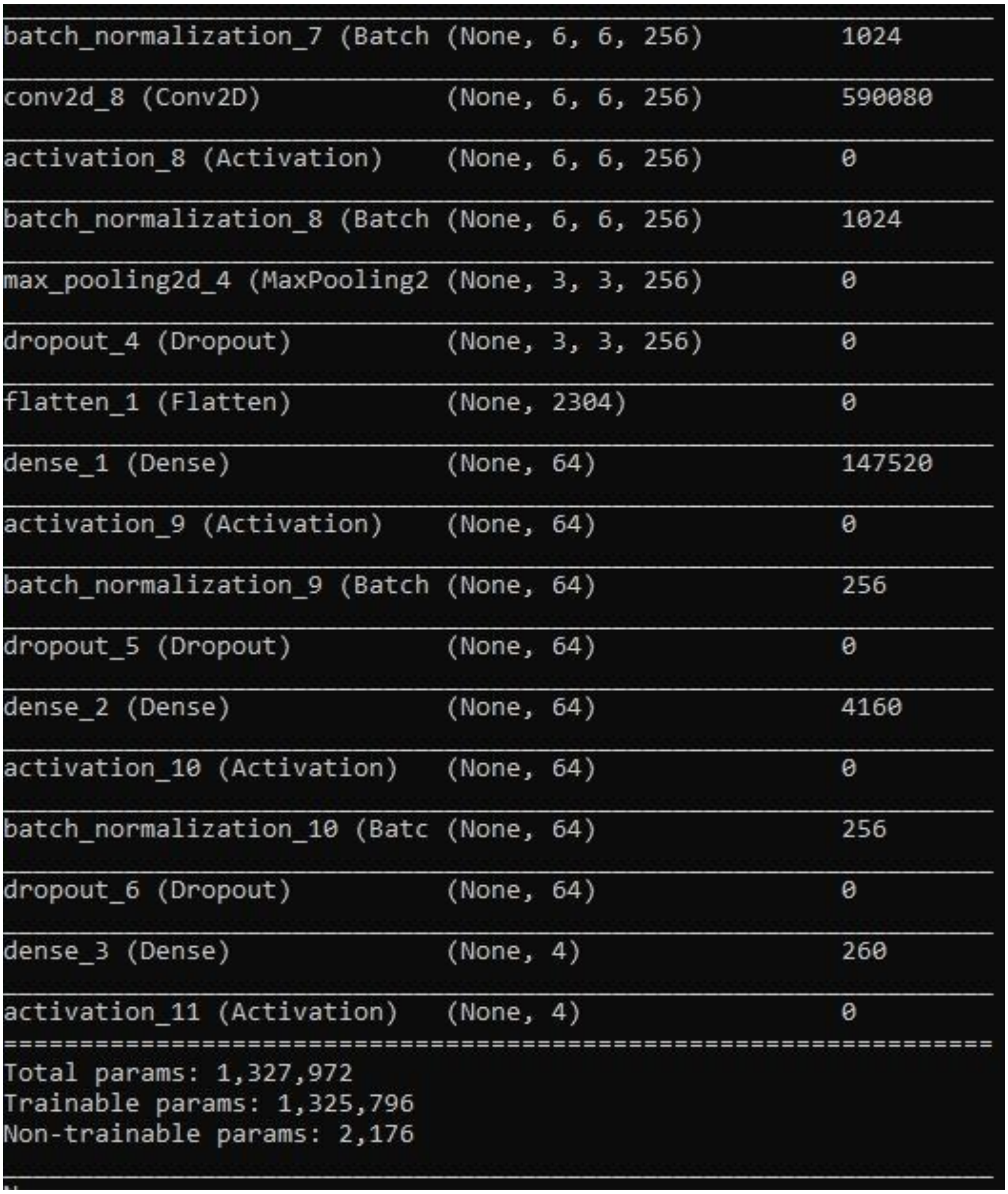
Batch Normalization: Batch normalization is used to stabilize and perhaps accelerate the learning process. It does so by applying a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1.

MaxPooling2D: Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

Dropout: Dropout is a technique used to prevent a model from overfitting. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase.



```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 48, 48, 32)        320
_________________________________________________________________
activation_1 (Activation)    (None, 48, 48, 32)        0
_________________________________________________________________
batch_normalization_1 (Batch (None, 48, 48, 32)        128
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 48, 48, 32)        9248
_________________________________________________________________
activation_2 (Activation)    (None, 48, 48, 32)        0
_________________________________________________________________
batch_normalization_2 (Batch (None, 48, 48, 32)        128
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 24, 24, 32)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 24, 24, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 24, 24, 64)        18496
_________________________________________________________________
activation_3 (Activation)    (None, 24, 24, 64)        0
_________________________________________________________________
batch_normalization_3 (Batch (None, 24, 24, 64)        256
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 24, 24, 64)        36928
_________________________________________________________________
activation_4 (Activation)    (None, 24, 24, 64)        0
_________________________________________________________________
batch_normalization_4 (Batch (None, 24, 24, 64)        256
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 12, 12, 64)        0
_________________________________________________________________
dropout_2 (Dropout)          (None, 12, 12, 64)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 12, 12, 128)       73856
_________________________________________________________________
activation_5 (Activation)    (None, 12, 12, 128)       0
_________________________________________________________________
batch_normalization_5 (Batch (None, 12, 12, 128)       512
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 12, 12, 128)       147584
_________________________________________________________________
activation_6 (Activation)    (None, 12, 12, 128)       0
_________________________________________________________________
batch_normalization_6 (Batch (None, 12, 12, 128)       512
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
dropout_3 (Dropout)          (None, 6, 6, 128)         0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 6, 6, 256)         295168
_________________________________________________________________
activation_7 (Activation)    (None, 6, 6, 256)         0
_________________________________________________________________
batch_normalization_7 (Batch (None, 6, 6, 256)         1024
_________________________________________________________________
```



```
________________________________________________________________
batch_normalization_7 (Batch (None, 6, 6, 256)         1024
________________________________________________________________
conv2d_8 (Conv2D)            (None, 6, 6, 256)         590080
________________________________________________________________
activation_8 (Activation)    (None, 6, 6, 256)         0
________________________________________________________________
batch_normalization_8 (Batch (None, 6, 6, 256)         1024
________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 3, 3, 256)         0
________________________________________________________________
dropout_4 (Dropout)          (None, 3, 3, 256)         0
________________________________________________________________
flatten_1 (Flatten)          (None, 2304)              0
________________________________________________________________
dense_1 (Dense)              (None, 64)                147520
________________________________________________________________
activation_9 (Activation)    (None, 64)                0
________________________________________________________________
batch_normalization_9 (Batch (None, 64)                256
________________________________________________________________
dropout_5 (Dropout)          (None, 64)                0
________________________________________________________________
dense_2 (Dense)              (None, 64)                4160
________________________________________________________________
activation_10 (Activation)   (None, 64)                0
________________________________________________________________
batch_normalization_10 (Batc (None, 64)                256
________________________________________________________________
dropout_6 (Dropout)          (None, 64)                0
________________________________________________________________
dense_3 (Dense)              (None, 4)                 260
________________________________________________________________
activation_11 (Activation)   (None, 4)                 0
================================================================
Total params: 1,327,972
Trainable params: 1,325,796
Non-trainable params: 2,176

________________________________________________________________
```

Figure 3.9: Model Summary

4. **Callback** - A callback is an object that can perform actions at various stages of training (e.g. at the start or end of an epoch, before or after a single batch, etc). For our classifier we have used ModelCheckpoint callback, which is used in conjunction with training using model.fit() to save a model or weights at some interval, so the model or weights can be loaded later to continue the training from the state saved.
5. **Early Stopping -** Now in order to halt the training process at real time we have used Early stopping method, that allows us to specify an arbitrary large

number of training epochs and stop training once the model performance stops improving on a hold out validation dataset.

6. **ReduceLROnPlateau** - Models often benefit from reducing the learning rate by a factor of 2-10 once learning stagnates. This callback monitors a quantity and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced.
7. **The Keras fit_generator function:**

Here, the fit_generator is used prior to generating our model because -

- Real-world datasets are often too large to fit into memory.
- They also tend to be challenging, requiring us to perform data augmentation to avoid overfitting and increase the ability of our model to generalize.

Hence we have used the fit_generator function to achieve the above.

3.2.4.3 Testing and Classification of emotion

After the classifier is made, before testing it with the facial input of the user we try to calculate the accuracy of our model by testing it with a validation set.

This verification is done because in the validation phase, we have used a part of the FER dataset itself to compute the accuracy. The FER dataset contains a total of 21,055 images, out of which 4005 belong to Anger, 4965 belong to Neutral, 7255 belong to Happy, 4830 belong to Sadness. Now, the part of the FER dataset that we took for the validation phase contains a total of 2590 images, out of which 491 belong to Anger, 879 belong to Happy, 626 belong to Neutral and 594 belong to Sadness.

Furthermore, the FER dataset contains various inconsistencies such as pictures of sketches, pictures with watermarks, pictures that do not fit with the particular emotion, etc. which leads to overfitting of our model. Hence, making it difficult to find the proper accuracy of our model.

After finding the accuracy of the classifier is obtained, our model can now be tested with the facial input in real time (the output of the Facial Image region Localization) and thereby, classifying the current emotion of the user.

3.2.4.4 Activity Suggestion

After the facial image is taken and an emotion is detected, the system now suggests a few activities based on the emotion which is detected. So in order to do that, the label is passed through a function which then proceeds to a list of suggested activities. This is done in an audio visual method. i.e- Text on screen and audio on screen so that the system can be used by anyone with visual impairment as well.

Our system now speaks/suggests the activities as follows:

- **For Happy & Neutral Emotions:**

  For “Happy” and “Neutral”, we have a selected few activities which we think would help sustain the happy emotion. This comprises opening popular websites such as Youtube, Google, Wikipedia, listening to an audiobook, asking for a movie recommendation, suggesting Exercises, playing music and getting today's news.

- **For Sad:**

  For “Sad”, we have a selected few activities which we think would help make a sad person feel better. This consists of telling a randomly generated joke to the user (through the use of an API), listening to an audiobook (suited for a person feeling sad), asking for a movie recommendation, suggesting Exercises and playing appropriate music.

- **For Anger:**

  For “Anger”, we first take a heartbeat reading from the user to estimate the BPM, Then we suggest a quick calm exercise that can help make the user feel at ease and take another heartbeat reading to check if the user has calmed down or not. After that, we give a list of activities that can be done after the said quick calm exercise. These activities include playing music, suggesting Exercises, watching a funny video, listening to an audiobook (suited for a person feeling Angry) and asking for a movie recommendation.

Now, the intricacies of the above activities are further explained as follows:

1. **HeartRate detection :**

Heart Rate can be estimated by imperceptible motions of the head caused by blood flow or by observing the blood oxygenated haemoglobin which absorbs the green light and measures the heart rate by detecting the variation of the green color intensity of the person's face.

The method that we tried to use is by considering the time series of the color component's value at any spatial location (pixel) and amplifying the changes in a given temporal frequency band of interest. Let us consider the instance, we automatically capture, and then try to amplify, a band of temporal frequencies that consists of plausible human heart-rates. The amplification reveals the changes in variation of greenness or redness as blood flows through the facial region. For our case, we try to study and then amplify the variation of pixel values over time, in a spatially-multiscale manner. In our Eulerian approach to motion magnification, we do not explicitly estimate motion, but rather exaggerate motion by amplifying temporal color changes at fixed positions.The approach that was used was to observe the series of values of colour in time scale at any pixel region (spatial location), and then amplify the change in a certain band of temporal-frequency. [19]

Fig. 3.10 shows the main framework of Eulerian video magnification. The video is processed frame by frame and the minute variations in the video are emphasized by the spatial and temporal processing. It should be noted that these minute variations cannot be detected by the naked eye. This involves several steps and can be described as follows:

1. Decomposition of the standard video sequence into different bands of spatial frequency and the subsequently applying a full Laplacian pyramid.
2. On each of the spatial bands, performing temporal processing. In this phase, according to the frequency band of interest, different types of band-pass filter would be chosen. And the temporal filtering approach can not only amplify the color variation, but also amplify the low amplitude motion as well.
3. Employing α, a magnification factor to multiply the extracted band-pass signal. Different applications specify the value of α and the boundary of α is

determined by δ, which is the image structure spatial wavelength and λ, which is the image structure spatial wavelength . The equation is given by

$$(1 + \alpha) * \delta < \lambda/8$$

4. Selecting a spatial frequency for cut-off. α is attenuated beyond this frequency.
5. Reconstruction of tReconstruct the signal by adding the magnified signal to the original one.

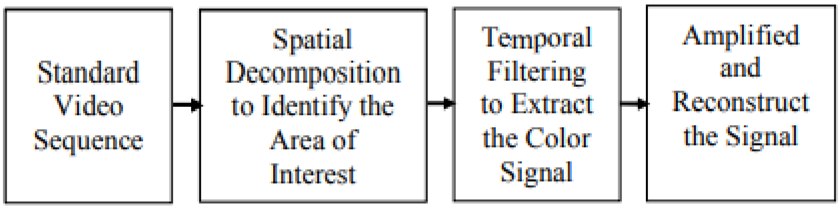

Standard Video Sequence
Spatial Decomposition to Identify the Area of Interest
Temporal Filtering to Extract the Color Signal
Amplified and Reconstruct the Signal

Figure 3.10: Framework of Eulerian Magnification.

Source: Using Eulerian Video Magnification Framework to Measure Pulse Transit Time Xiaochuan He, Rafik A. Goubran, IEEE Fellow, Xiaoping P. Liu IEEE Senior Member Carleton University: Department of Systems and Computer Engineering Ottawa, Canada xche@sce.carleton.ca

2. **Movie Recommendation :**

For implementing a movie recommender system, we have the option of either using the "trending list" concept where we could recommend by using the "what's popular list "or we can even use a collaborative filtering approach which does not need anything else except the user's interactions and feedback.

However, as we are trying to deal with a single particular user's preferences so we have another approach which is the Content based filtering. Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.

- **Content-based filtering**

Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.

The following figure shows a feature matrix where each row represents an app and each column represents a feature. Features could include categories (such as Education, Casual, Health), the publisher of the app, and many others. To simplify, we assume that this feature matrix is binary: a non-zero value means the app has that feature.

We can also represent the user in the same feature space. Some of the user-related features could be explicitly provided by the user. For example, a user selects "Entertainment apps" in their profile. Other features can be implicit, based on the apps they have previously installed. For example, the user installed another app published by Science R Us.

The model should recommend items relevant to this user. To do so, the user must first pick a similarity metric (for example, dot product). Then, he/she must set up the system to score each candidate item according to this similarity metric. We should note that the recommendations are specific to this user, as the model did not use any information about other users.[25]

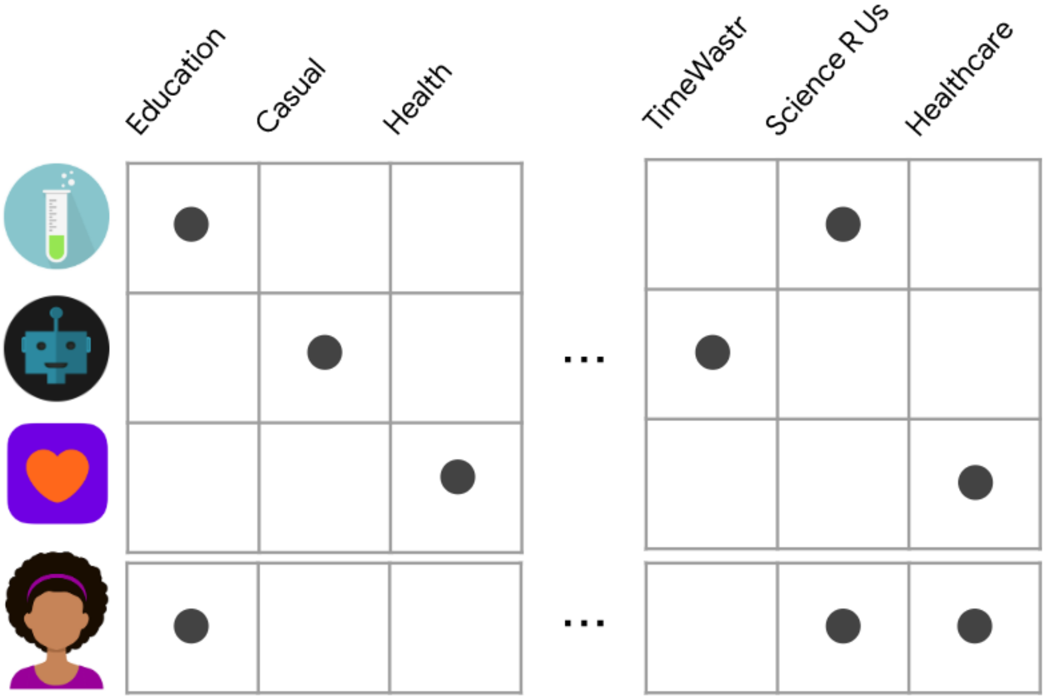
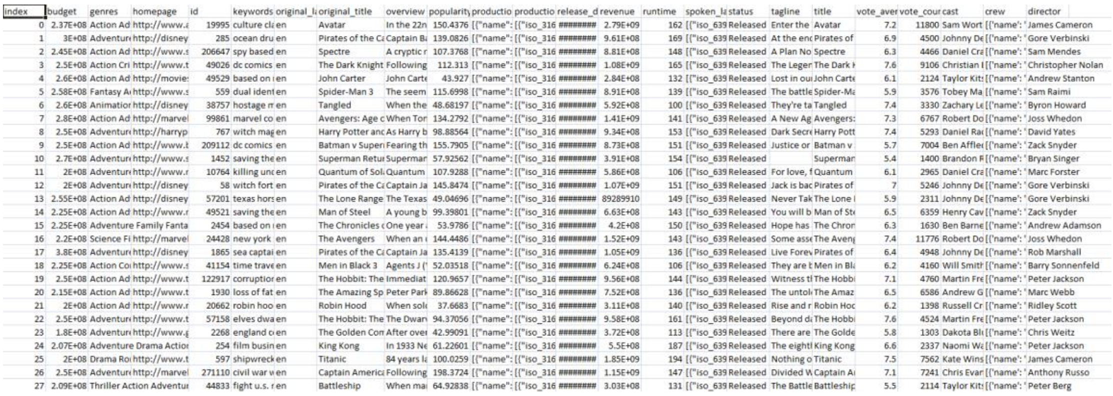


Figure 3.11: Content-based Filtering

Source:https://developers.google.com/machine-learning/recommendation/content-based/basics

So for recommending movies we have used content based filtering using scikit learn. We have used a dataset that contains the description of 4802 different movies taken from IMDB (Internet Movie Database), each movie in the dataset is described by 24 different attributes like budget, genres, homepage, overview, popularity, votes, etc.



index	budget	genres	homepage	id	keywords	original_l	original_title	overview	popularity	productio	productio	release_d	revenue	runtime	spoken_la	status	tagline	title	vote_aver	vote_cour	cast	crew	director
0	2.37E+08	Action Ad	http://www.a	19995	culture cla	en	Avatar	In the 22n	150.4376	[{"name":	[{"iso_316	########	2.79E+09	162	[{"iso_639	Released	Enter the	Avatar	7.2	11800	Sam Wort	[{'name': '	James Cameron
1	3E+08	Adventur	http://disney	285	ocean dru	en	Pirates of the Ca	Captain Ba	139.0826	[{"name":	[{"iso_316	########	9.61E+08	169	[{"iso_639	Released	At the enc	Pirates of	6.9	4500	Johnny De	[{'name': '	Gore Verbinski
2	2.45E+08	Action Ad	http://www.s	206647	spy based	en	Spectre	A cryptic r	107.3768	[{"name":	[{"iso_316	########	8.81E+08	148	[{"iso_639	Released	A Plan No	Spectre	6.3	4466	Daniel Cra	[{'name': '	Sam Mendes
3	2.5E+08	Action Cri	http://www.t	49026	dc comics	en	The Dark Knight	Following	112.313	[{"name":	[{"iso_316	########	1.08E+09	165	[{"iso_639	Released	The Legen	The Dark K	7.6	9106	Christian B	[{'name': '	Christopher Nolan
4	2.6E+08	Action Ad	http://movies	49529	based on	en	John Carter	John Carte	43.927	[{"name":	[{"iso_316	########	2.84E+08	132	[{"iso_639	Released	Lost in oui	John Carte	6.1	2124	Taylor Kits	[{'name': '	Andrew Stanton
5	2.58E+08	Fantasy A	http://www.s	559	dual ident	en	Spider-Man 3	The seem	115.6998	[{"name":	[{"iso_316	########	8.91E+08	139	[{"iso_639	Released	The battle	Spider-Ma	5.9	3576	Tobey Ma	[{'name': '	Sam Raimi
6	2.6E+08	Animation	http://disney	38757	hostage m	en	Tangled	When the	48.68197	[{"name":	[{"iso_316	########	5.92E+08	100	[{"iso_639	Released	They're ta	Tangled	7.4	3330	Zachary Le	[{'name': '	Byron Howard
7	2.8E+08	Action Ad	http://marvel	99861	marvel co	en	Avengers: Age c	When Tor	134.2792	[{"name":	[{"iso_316	########	1.41E+09	141	[{"iso_639	Released	A New Ag	Avengers:	7.3	6767	Robert Do	[{'name': '	Joss Whedon
8	2.5E+08	Adventur	http://harryp	767	witch mag	en	Harry Potter anc	As Harry b	98.88564	[{"name":	[{"iso_316	########	9.34E+08	153	[{"iso_639	Released	Dark Secre	Harry Pott	7.4	5293	Daniel Rac	[{'name': '	David Yates
9	2.5E+08	Action Ad	http://www.b	209112	dc comics	en	Batman v Super	Fearing th	155.7905	[{"name":	[{"iso_316	########	8.73E+08	151	[{"iso_639	Released	Justice or	Batman v	5.7	7004	Ben Affle	[{'name': '	Zack Snyder
10	2.7E+08	Adventur	http://www.s	1452	saving the	en	Superman Retur	Superman	57.92562	[{"name":	[{"iso_316	########	3.91E+08	154	[{"iso_639	Released		Superman	5.4	1400	Brandon R	[{'name': '	Bryan Singer
11	2E+08	Adventur	http://www.r	10764	killing unc	en	Quantum of Sol	Quantum	107.9288	[{"name":	[{"iso_316	########	5.86E+08	106	[{"iso_639	Released	For love, f	Quantum	6.1	2965	Daniel Cra	[{'name': '	Marc Forster
12	2E+08	Adventur	http://disney	58	witch fort	en	Pirates of the Ca	Captain Ja	145.8474	[{"name":	[{"iso_316	########	1.07E+09	151	[{"iso_639	Released	Jack is bac	Pirates of	7	5246	Johnny De	[{'name': '	Gore Verbinski
13	2.55E+08	Action Ad	http://disney	57201	texas hors	en	The Lone Range	The Texas	49.04696	[{"name":	[{"iso_316	########	89289910	149	[{"iso_639	Released	Never Tak	The Lone	5.9	2311	Johnny De	[{'name': '	Gore Verbinski
14	2.25E+08	Action Ad	http://www.r	49521	saving the	en	Man of Steel	A young b	99.39801	[{"name":	[{"iso_316	########	6.63E+08	143	[{"iso_639	Released	You will b	Man of Ste	6.5	6359	Henry Cav	[{'name': '	Zack Snyder
15	2.25E+08	Adventure Family Fanta		2454	based on	en	The Chronicles c	One year	53.9786	[{"name":	[{"iso_316	########	4.2E+08	150	[{"iso_639	Released	Hope has	The Chron	6.3	1630	Ben Barne	[{'name': '	Andrew Adamson
16	2.2E+08	Science Fi	http://marvel	24428	new york	en	The Avengers	When an	144.4486	[{"name":	[{"iso_316	########	1.52E+09	143	[{"iso_639	Released	Some asse	The Aveng	7.4	11776	Robert Do	[{'name': '	Joss Whedon
17	3.8E+08	Adventur	http://disney	1865	sea captai	en	Pirates of the Ca	Captain Ja	135.4139	[{"name":	[{"iso_316	########	1.05E+09	136	[{"iso_639	Released	Live Forev	Pirates of	6.4	4948	Johnny De	[{'name': '	Rob Marshall
18	2.25E+08	Action Co	http://www.s	41154	time trave	en	Men in Black 3	Agents J ('	52.03518	[{"name":	[{"iso_316	########	6.24E+08	106	[{"iso_639	Released	They are t	Men in Bla	6.2	4160	Will Smith	[{'name': '	Barry Sonnenfeld
19	2.5E+08	Action Ad	http://www.t	122917	corruption	en	The Hobbit: The	Immediat	120.9657	[{"name":	[{"iso_316	########	9.56E+08	144	[{"iso_639	Released	Witness th	The Hobbi	7.1	4760	Martin Fre	[{'name': '	Peter Jackson
20	2.15E+08	Action Ad	http://www.t	1930	loss of fat	en	The Amazing Sp	Peter Park	89.86628	[{"name":	[{"iso_316	########	7.52E+08	136	[{"iso_639	Released	The untol	The Amaz	6.5	6586	Andrew G	[{'name': '	Marc Webb
21	2E+08	Action Ad	http://www.r	20662	robin hoo	en	Robin Hood	When sol	37.6683	[{"name":	[{"iso_316	########	3.11E+08	140	[{"iso_639	Released	Rise and r	Robin Hoc	6.2	1398	Russell Cr	[{'name': '	Ridley Scott
22	2.5E+08	Adventur	http://www.t	57158	elves dwa	en	The Hobbit: The	The Dwar	94.37056	[{"name":	[{"iso_316	########	9.58E+08	161	[{"iso_639	Released	Beyond da	The Hobbi	7.6	4524	Martin Fre	[{'name': '	Peter Jackson
23	1.8E+08	Adventur	http://www.g	2268	england c	en	The Golden Con	After over	42.99091	[{"name":	[{"iso_316	########	3.72E+08	113	[{"iso_639	Released	There are	The Golde	5.8	1303	Dakota Bl	[{'name': '	Chris Weitz
24	2.07E+08	Adventure Drama Actio		254	film busin	en	King Kong	In 1933 Ne	61.22601	[{"name":	[{"iso_316	########	5.5E+08	187	[{"iso_639	Released	The eightl	King Kong	6.6	2337	Naomi Wa	[{'name': '	Peter Jackson
25	2E+08	Drama Ro	http://www.t	597	shipwreck	en	Titanic	84 years la	100.0259	[{"name":	[{"iso_316	########	1.85E+09	194	[{"iso_639	Released	Nothing o	Titanic	7.5	7562	Kate Wins	[{'name': '	James Cameron
26	2.5E+08	Adventur	http://marvel	271110	civil war w	en	Captain Americ	Following	198.3724	[{"name":	[{"iso_316	########	1.15E+09	147	[{"iso_639	Released	Divided W	Captain A	7.1	7241	Chris Evar	[{'name': '	Anthony Russo
27	2.09E+08	Thriller Action Adventur		44833	fight u.s. r	en	Battleship	When ma	64.92838	[{"name":	[{"iso_316	########	3.03E+08	131	[{"iso_639	Released	The Battle	Battleship	5.5	2114	Taylor Kits	[{'name': '	Peter Berg

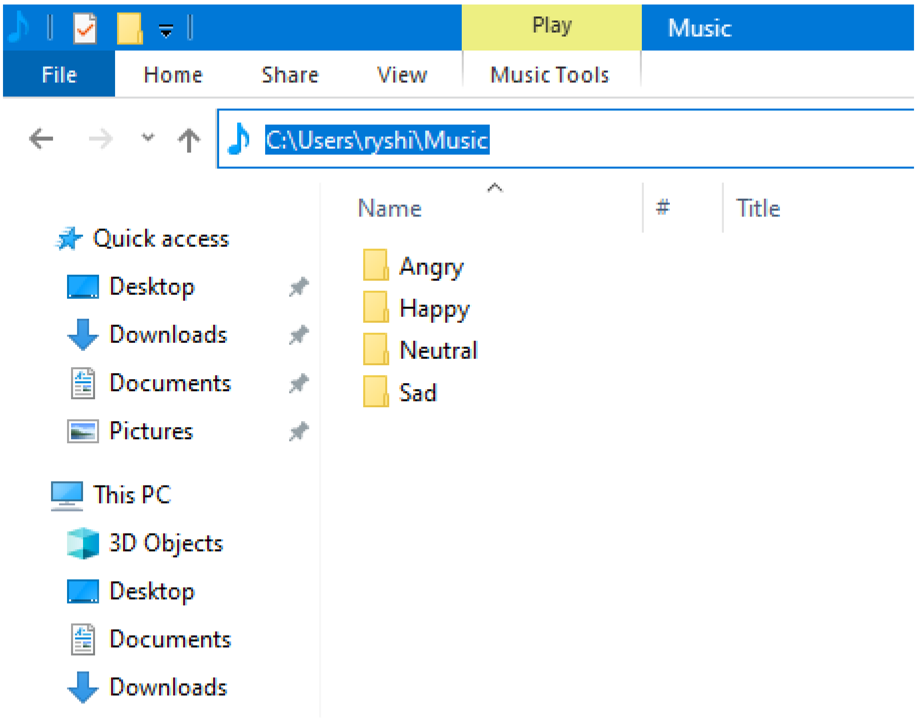
Figure 3.12: The movie dataset

For the purpose of finding similarities between the movies, we have used libraries like the count vectorizer that convert the collection of text documents to a matrix of token counts and cosine similarity which helps us to calculate the similarity scores between the movies. And based on the similarity scores, similar movies can be recommended to the user.

3. **Audio playback :**

This is done by calling the OS function to open the system default app for playing music files. The folders are divided based on the labels of the emotions and the emotion of the person determines which song is currently going to be played.

The folder is shown as below:



C:\Users\ryshi\Music

Figure 3.13: Folder representation

4. **GoogleNews API :**

This API enables EBASS to gather News data depending on the query keyword that can be said by the user to search news about.

5. **Jokes API :**

The API randomly tells the user a one liner joke.

6. **Voice input & Output :**

EBASS listens for voice commands while narrating out the on screen text to help assist visually impaired people.

Chapter 4 - System Design

4.1 System Architechture

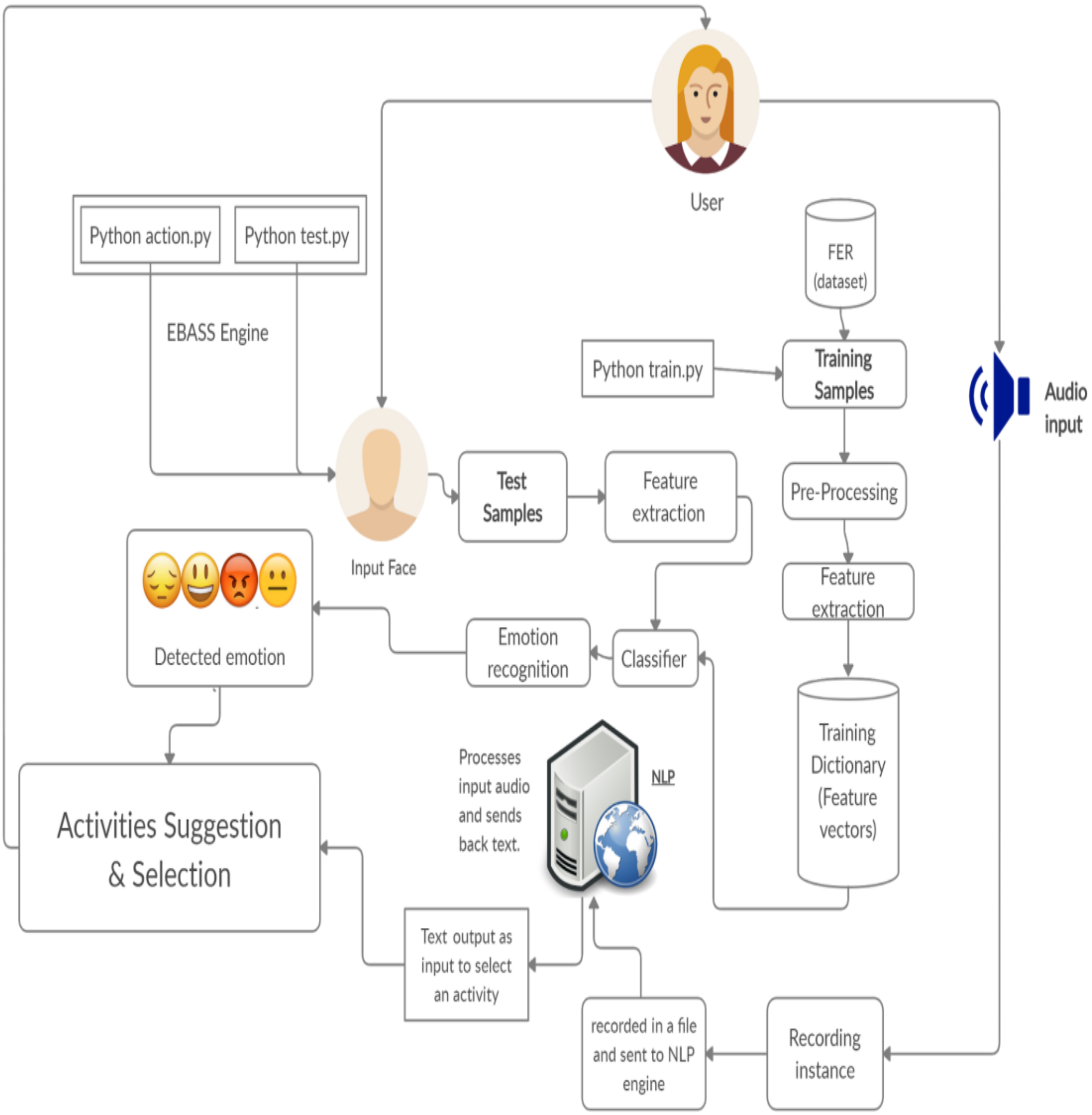


Figure 4.1: System architechture

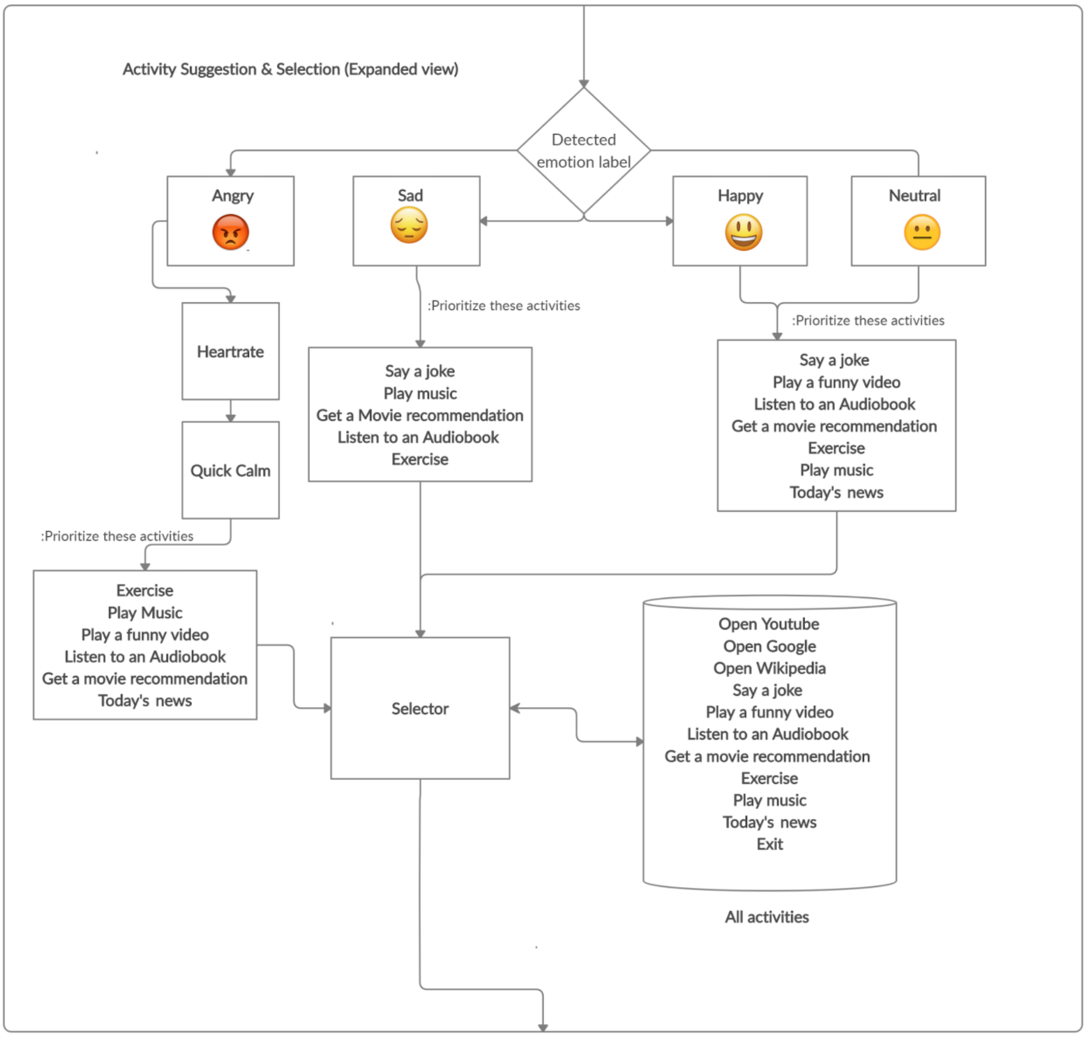


Fig 4.2: Expanded view of Activity Suggestion & selection

The above system architecture is the expanded view of the “Activities Suggestion and Selection” phase in the original pipeline shown in fig 4.2.

Figures 4.1 and 4.2 show the system architecture of our system, it shows how we have connected different modules to integrate our system. The classifier that classifies the emotions of the user is the outcome of the module train.py. In order to build this classifier we have used the FER 2013 dataset that contains 21,055 images of four different emotions which we had depicted in the earlier chapter. For training our classifier using this dataset we have used CNN for extracting the features. Under CNN we have used the VGG 16 network architecture. So using the Python train.py we train our classifier and pre process the samples, upon feature extraction we can get the defined feature vectors which defines the training dictionary and hence results in our classifier.

Now the python action.py module comes into action where the python test.py is embedded in it. Here we take the user's face as input by real time processing. The facial input is treated as a test sample to our classifier and upon detecting the facial region and comparing with the training dictionary our classifier recognizes the current emotion of the user. As the emotion is recognized, based on that particular emotion a certain list of activities is suggested by our system. But if we detect the emotion as "Angry", then we first take a heartbeat reading from the user to estimate the BPM, Then we suggest a quick calm exercise that can help make the user feel at ease and take another heartbeat reading to check if the user has calmed down or not. Then we suggest the activities.

At this point of time, upon looking into the activity list we will take the command from the user through an audio input. This is done by using the pyttsx library which includes the Speech Application Programming Interface. This audio input instance is recorded in a file and sent to the Natural Language Processing engine which processes the input audio and sends the text data which is treated as the command to select an appropriate activity. After the audio input is recognized, the activity selector compares the input with the activity list and selects that particular activity as commanded by the user.

4.2 Software Design

4.2.1 Use of major Python libraries/utilities

1. OpenCV

(Open Source Computer Vision Library) is an open source computer vision and machine learning software library. It was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and machine learning algorithms.

2. NumPy

Numpy is a library for the Python Programming Language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

3. Time

Python has defined a module, “time” which allows us to handle various operations regarding time, its conversions and representations, which find its use in various applications in life. The beginning of time is started measuring from 1 January, 12:00 am, 1970 and this very time is termed as “epoch” in Python.

4. Keras

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow. We use it for Dense,Dropout,Activation,Flatten,BatchNormalization,Conv2D,MaxPooling2D, And other preprocessing on images. We also use RMSprop,SGD,Adam as optimizers and ModelCheckpoint, EarlyStopping, ReduceLROnPlateau as callback methods.

5. TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. We are using the GPU enabled version of Tensorflow and with the help of Keras we are able to do the functions required.

6. Pyttsx3 API

pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline, and is compatible with both Python 2 and 3.

The pyttsx3 module supports two voices, first is female and the second is male which is provided by “sapi5” for windows.

It supports three TTS engines :

- sapi5 – SAPI5 on Windows
- nsss – NSSpeechSynthesizer on Mac OS X
- espeak – eSpeak on every other platform

7. SpeechRecognition 3.8.1

Library for performing speech recognition, with support for several engines and APIs, online and offline.

8. PyAudio

PyAudio provides Python bindings for PortAudio, the cross-platform audio I/O library. With PyAudio, you can easily use Python to play and record audio on a variety of platforms.

9. GoogleNews API

The GoogleNews API provides news and search capabilities from news.google.com.

10. axju-jokes API

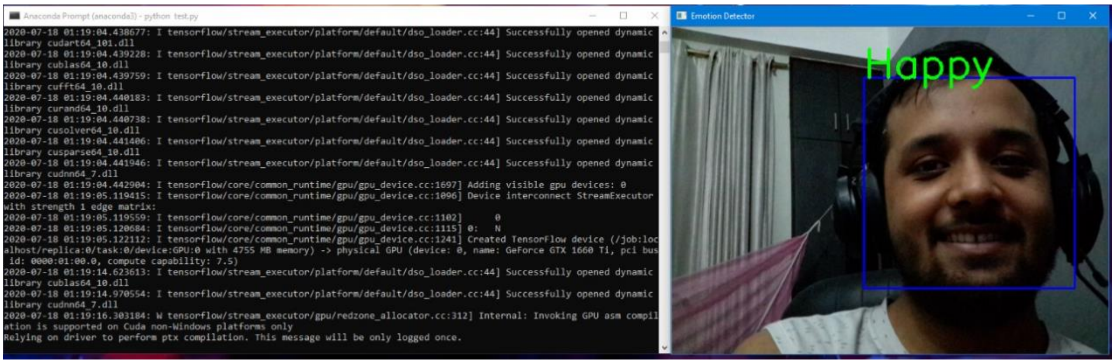
An API that returns random one-liner jokes.

Chapter 5 - Results and Discussion

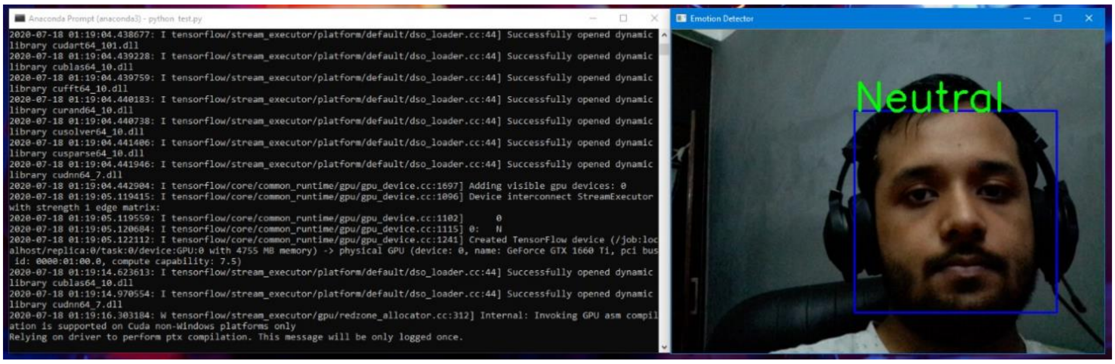
In this chapter, we test the system and show you the screenshots that

Our system was able to properly detect the 4 emotions (namely “Happy”, “Neutral”, “Sad” and “Angry”) on a real-time video feed.

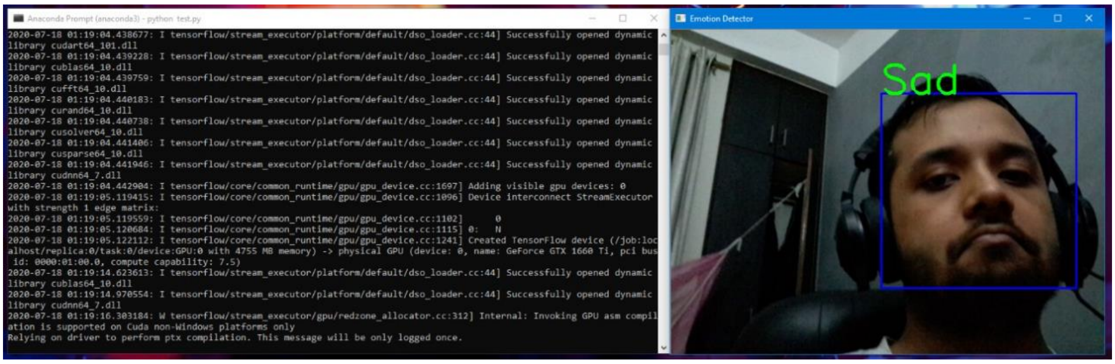
Below are the attached screenshots which show the detection of the emotion.



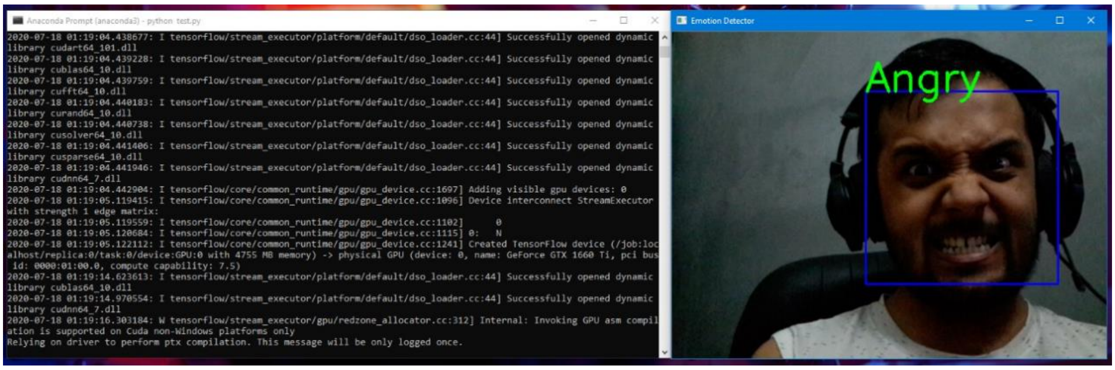
```
2020-07-18 01:19:04.438677: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudart64_101.dll
2020-07-18 01:19:04.439228: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:04.439759: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cufft64_10.dll
2020-07-18 01:19:04.440183: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library curand64_10.dll
2020-07-18 01:19:04.440738: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusolver64_10.dll
2020-07-18 01:19:04.441406: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusparse64_10.dll
2020-07-18 01:19:04.441946: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:04.442904: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1697] Adding visible gpu devices: 0
2020-07-18 01:19:05.119415: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1096] Device interconnect StreamExecutor
with strength 1 edge matrix:
2020-07-18 01:19:05.119559: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102]      0
2020-07-18 01:19:05.120684: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1115] 0:   N
2020-07-18 01:19:05.122112: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:loc
alhost/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus
 id: 0000:01:00.0, compute capability: 7.5)
2020-07-18 01:19:14.623613: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:14.970554: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:16.303184: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compil
ation is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
```



```
2020-07-18 01:19:04.438677: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudart64_101.dll
2020-07-18 01:19:04.439228: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:04.439759: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cufft64_10.dll
2020-07-18 01:19:04.440183: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library curand64_10.dll
2020-07-18 01:19:04.440738: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusolver64_10.dll
2020-07-18 01:19:04.441406: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusparse64_10.dll
2020-07-18 01:19:04.441946: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:04.442904: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1697] Adding visible gpu devices: 0
2020-07-18 01:19:05.119415: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1096] Device interconnect StreamExecutor
with strength 1 edge matrix:
2020-07-18 01:19:05.119559: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102]      0
2020-07-18 01:19:05.120684: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1115] 0:   N
2020-07-18 01:19:05.122112: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:loc
alhost/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus
 id: 0000:01:00.0, compute capability: 7.5)
2020-07-18 01:19:14.623613: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:14.970554: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:16.303184: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compil
ation is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
```



```
2020-07-18 01:19:04.438677: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudart64_101.dll
2020-07-18 01:19:04.439228: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:04.439759: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cufft64_10.dll
2020-07-18 01:19:04.440183: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library curand64_10.dll
2020-07-18 01:19:04.440738: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusolver64_10.dll
2020-07-18 01:19:04.441406: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusparse64_10.dll
2020-07-18 01:19:04.441946: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:04.442904: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1697] Adding visible gpu devices: 0
2020-07-18 01:19:05.119415: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1096] Device interconnect StreamExecutor
with strength 1 edge matrix:
2020-07-18 01:19:05.119559: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102]      0
2020-07-18 01:19:05.120684: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1115] 0:   N
2020-07-18 01:19:05.122112: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:loc
alhost/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus
 id: 0000:01:00.0, compute capability: 7.5)
2020-07-18 01:19:14.623613: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:14.970554: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:16.303184: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compil
ation is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
```



```
2020-07-18 01:19:04.438677: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudart64_101.dll
2020-07-18 01:19:04.439228: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:04.439759: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cufft64_10.dll
2020-07-18 01:19:04.440183: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library curand64_10.dll
2020-07-18 01:19:04.440738: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusolver64_10.dll
2020-07-18 01:19:04.441406: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cusparse64_10.dll
2020-07-18 01:19:04.441946: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:04.442904: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1697] Adding visible gpu devices: 0
2020-07-18 01:19:05.119415: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1096] Device interconnect StreamExecutor
with strength 1 edge matrix:
2020-07-18 01:19:05.119559: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102]      0
2020-07-18 01:19:05.120684: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1115] 0:   N
2020-07-18 01:19:05.122112: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:loc
alhost/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus
 id: 0000:01:00.0, compute capability: 7.5)
2020-07-18 01:19:14.623613: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cublas64_10.dll
2020-07-18 01:19:14.970554: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic
library cudnn64_7.dll
2020-07-18 01:19:16.303184: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compil
ation is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
```

Figure 5.1: Emotion detection

As we are checking for the emotion in real-time, therefore the emotions may vary which may lead to incorrect emotion “classification”.

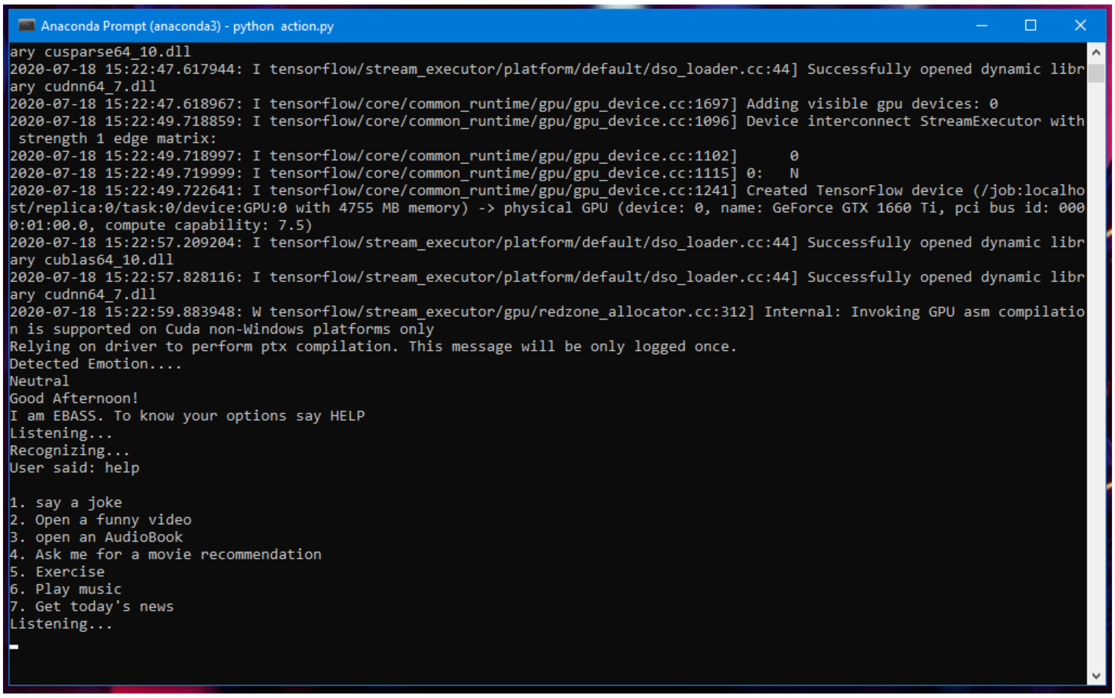
Moreover, the emotion detection is done on the frames of the live-video feed. Hence, technically speaking, each frame can have separate emotions.

Therefore, to fix these inconsistencies, we have inserted two conditions -

- The first one is where the processing for the feature extraction can be done for a certain amount of time.

  This means that the emotion classification will automatically close after the certain amount of time is over.
- We have also included a counter for the emotions detected in each frame. So, the dominating/most-prevalent emotion observed from all the frame counters is taken as the final emotion.

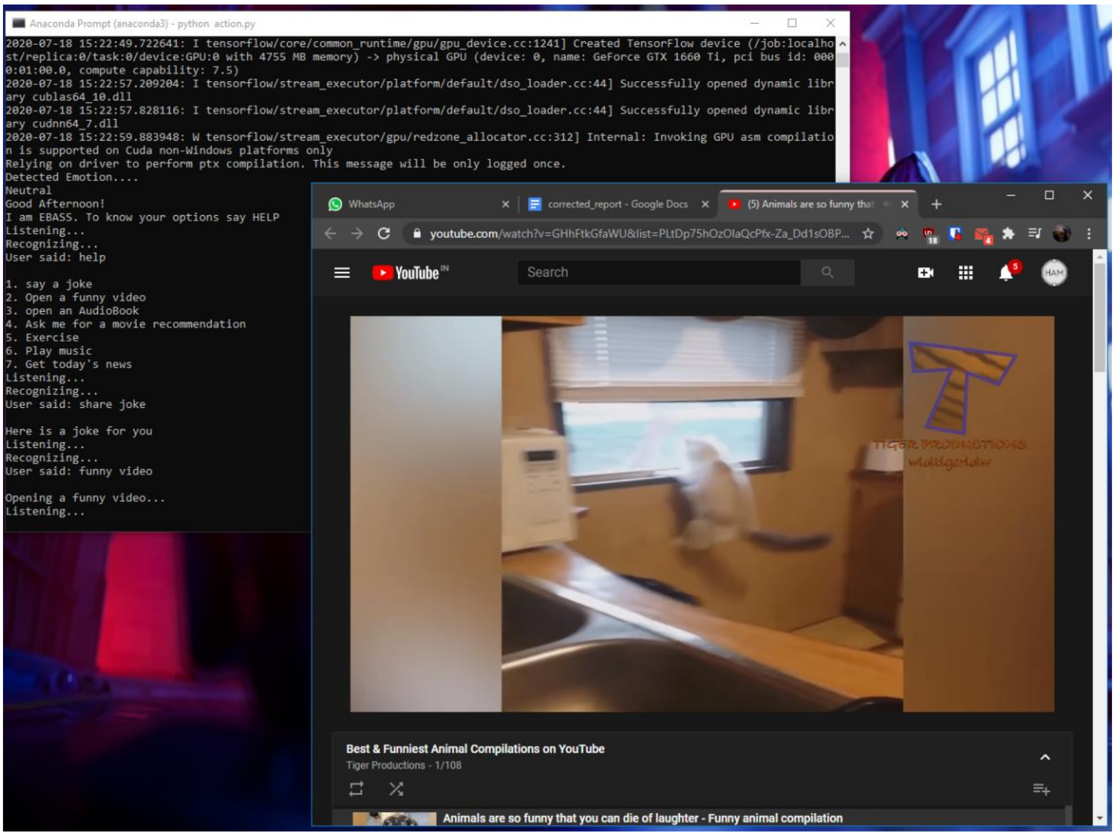
It should also be noted that the requirements must be satisfied for the proper detection of emotions. This includes having a proper working webcam on your computer, preferably of high resolution and good lighting conditions.



```
ary cusparse64_10.dll
2020-07-18 15:22:47.617944: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic libr
ary cudnn64_7.dll
2020-07-18 15:22:47.618967: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1697] Adding visible gpu devices: 0
2020-07-18 15:22:49.718859: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1096] Device interconnect StreamExecutor with
 strength 1 edge matrix:
2020-07-18 15:22:49.718997: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102]      0
2020-07-18 15:22:49.719999: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1115] 0:   N
2020-07-18 15:22:49.722641: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:localho
st/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus id: 000
0:01:00.0, compute capability: 7.5)
2020-07-18 15:22:57.209204: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic libr
ary cublas64_10.dll
2020-07-18 15:22:57.828116: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic libr
ary cudnn64_7.dll
2020-07-18 15:22:59.883948: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compilatio
n is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
Detected Emotion....
Neutral
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

1. say a joke
2. Open a funny video
3. open an AudioBook
4. Ask me for a movie recommendation
5. Exercise
6. Play music
7. Get today's news
Listening...
```

Figure 5.2: Activities for Neutral emotion.

Here Fig. 5.2 is a screenshot above showing the help file of Neutral Emotion.

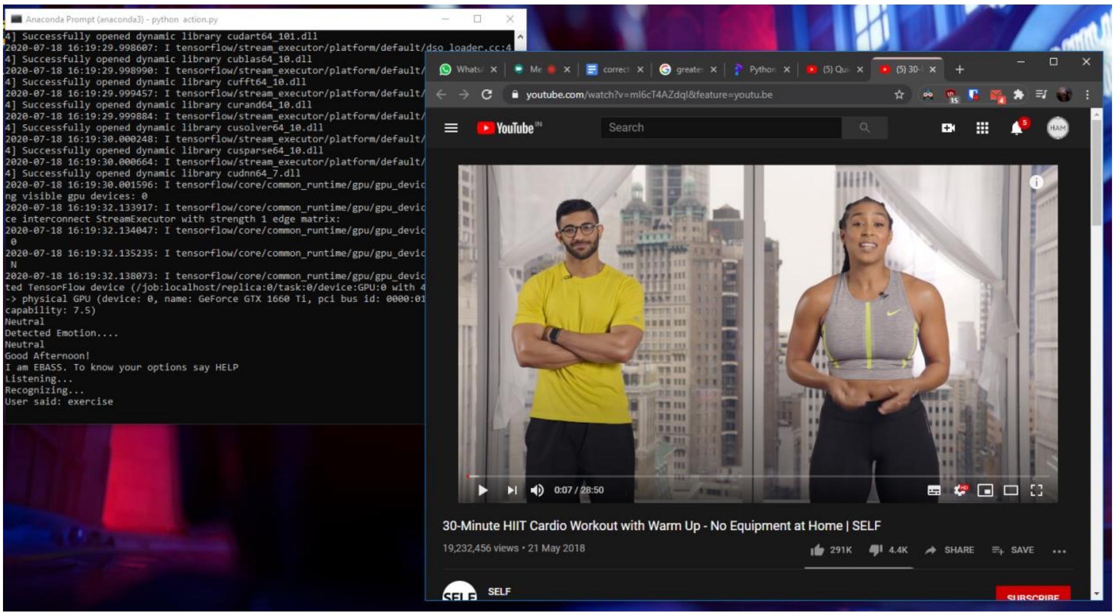


Anaconda Prompt (anaconda3) - python action.py
WhatsApp
corrected_report - Google Docs
(5) Animals are so funny that
youtube.com/watch?v=GHhFtkGfaWU&list=PLtDp75hOzOlaQcPfx-Za_Dd1sOBP...
YouTube IN
Search
Best & Funniest Animal Compilations on YouTube
Tiger Productions - 1/108
Animals are so funny that you can die of laughter - Funny animal compilation

```
2020-07-18 15:22:49.722641: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1241] Created TensorFlow device (/job:localho
st/replica:0/task:0/device:GPU:0 with 4755 MB memory) -> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus id: 000
0:01:00.0, compute capability: 7.5)
2020-07-18 15:22:57.209204: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic libr
ary cublas64_10.dll
2020-07-18 15:22:57.828116: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic libr
ary cudnn64_7.dll
2020-07-18 15:22:59.883948: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] Internal: Invoking GPU asm compilatio
n is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
Detected Emotion....
Neutral
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

1. say a joke
2. Open a funny video
3. open an AudioBook
4. Ask me for a movie recommendation
5. Exercise
6. Play music
7. Get today's news
Listening...
Recognizing...
User said: share joke

Here is a joke for you
Listening...
Recognizing...
User said: funny video

Opening a funny video...
Listening...
```

Figure 5.3: Funny video suggestion.

The above screenshot (Fig 5.3) shows a funny video being played to the user upon his query.

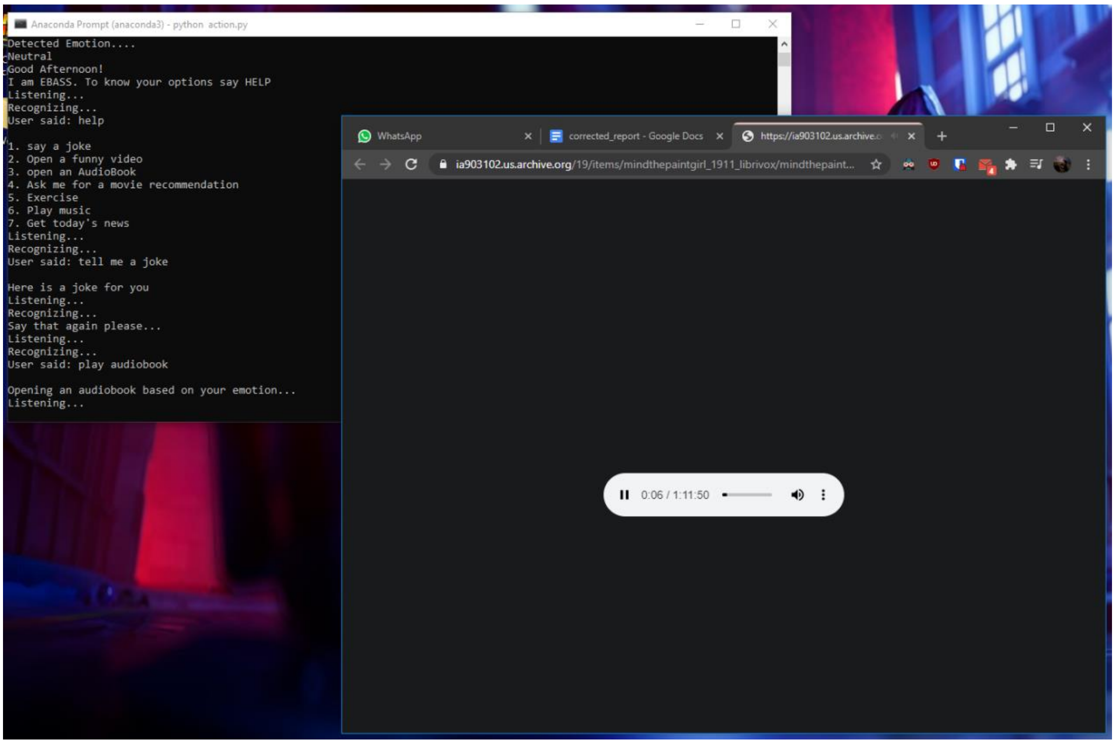


```
4] Successfully opened dynamic library cudart64_101.dll
2020-07-18 16:19:29.998607: I tensorflow/stream_executor/platform/default/dso_loader.cc:4
4] Successfully opened dynamic library cublas64_10.dll
2020-07-18 16:19:29.998990: I tensorflow/stream_executor/platform/default/
4] Successfully opened dynamic library cufft64_10.dll
2020-07-18 16:19:29.999457: I tensorflow/stream_executor/platform/default/
4] Successfully opened dynamic library curand64_10.dll
2020-07-18 16:19:29.999884: I tensorflow/stream_executor/platform/default/
4] Successfully opened dynamic library cusolver64_10.dll
2020-07-18 16:19:30.000248: I tensorflow/stream_executor/platform/default/
4] Successfully opened dynamic library cusparse64_10.dll
2020-07-18 16:19:30.000664: I tensorflow/stream_executor/platform/default/
4] Successfully opened dynamic library cudnn64_7.dll
2020-07-18 16:19:30.001596: I tensorflow/core/common_runtime/gpu/gpu_devic
ng visible gpu devices: 0
2020-07-18 16:19:32.133917: I tensorflow/core/common_runtime/gpu/gpu_devic
ce interconnect StreamExecutor with strength 1 edge matrix:
2020-07-18 16:19:32.134047: I tensorflow/core/common_runtime/gpu/gpu_devic
 0
2020-07-18 16:19:32.135235: I tensorflow/core/common_runtime/gpu/gpu_devic
 N
2020-07-18 16:19:32.138073: I tensorflow/core/common_runtime/gpu/gpu_devic
ted TensorFlow device (/job:localhost/replica:0/task:0/device:GPU:0 with 4
-> physical GPU (device: 0, name: GeForce GTX 1660 Ti, pci bus id: 0000:01
capability: 7.5)
Neutral
Detected Emotion....
Neutral
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: exercise
```

Anaconda Prompt (anaconda3) - python action.py
youtube.com/watch?v=ml6cT4AZdqI&feature=youtu.be
YouTube
Search
0:07 / 28:50
30-Minute HIIT Cardio Workout with Warm Up - No Equipment at Home | SELF
19,232,456 views • 21 May 2018
291K
4.4K
SHARE
SAVE
SELF
SUBSCRIBE

Figure 5.4: Exercise suggestion.

The above screenshot (Fig 5.4) shows an Exercise video being played to the user based upon his query.

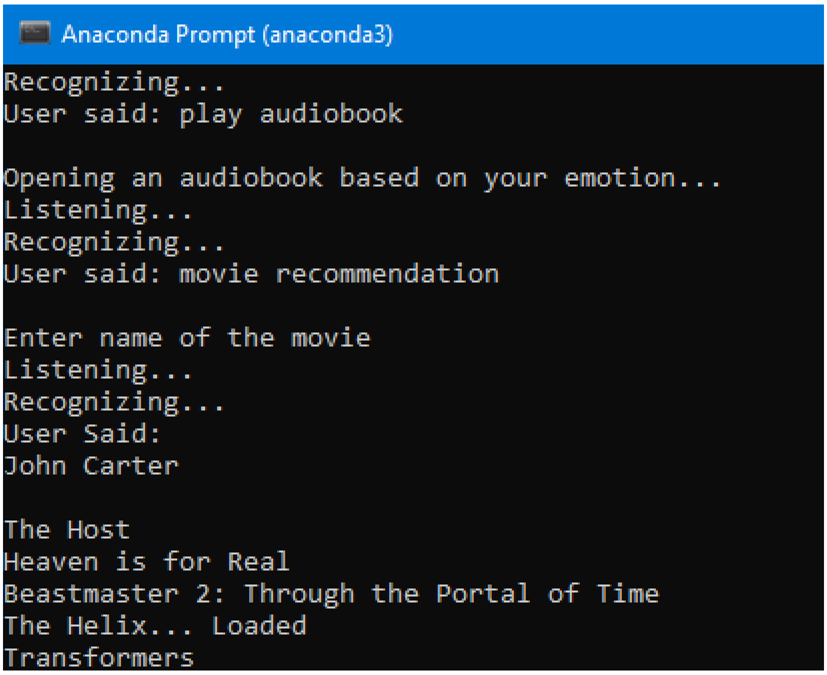


```
Detected Emotion....
Neutral
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

1. say a joke
2. Open a funny video
3. open an AudioBook
4. Ask me for a movie recommendation
5. Exercise
6. Play music
7. Get today's news
Listening...
Recognizing...
User said: tell me a joke

Here is a joke for you
Listening...
Recognizing...
Say that again please...
Listening...
Recognizing...
User said: play audiobook

Opening an audiobook based on your emotion...
Listening...
```

Anaconda Prompt (anaconda3) - python action.py
WhatsApp
corrected_report - Google Docs
https://ia903102.us.archive.o
ia903102.us.archive.org/19/items/mindthepaintgirl_1911_librivox/mindthepaint...
0:06 / 1:11:50

Figure 5.5: Audiobook playback.

The above screenshot (Fig 5.5) shows an audiobook being played to the user based upon both his mood and query.

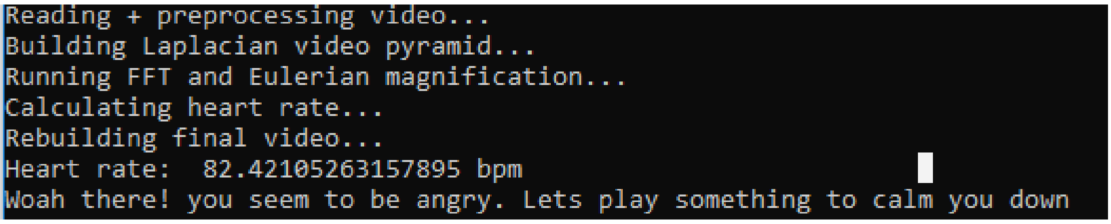


```
Recognizing...
User said: play audiobook

Opening an audiobook based on your emotion...
Listening...
Recognizing...
User said: movie recommendation

Enter name of the movie
Listening...
Recognizing...
User Said:
John Carter

The Host
Heaven is for Real
Beastmaster 2: Through the Portal of Time
The Helix... Loaded
Transformers
```

Figure 5.6: Movie recommendation.

It should be noted the csv file contains movies upto the time of creation of the dataset. This means that the movies that were released after the creation of the dataset are not taken into account. For e.g The system will not be able to recommend movies for “Extraction” since it is not included in the csv file yet. Hence, updation of the csv file is necessary for accurate movie recommendation.

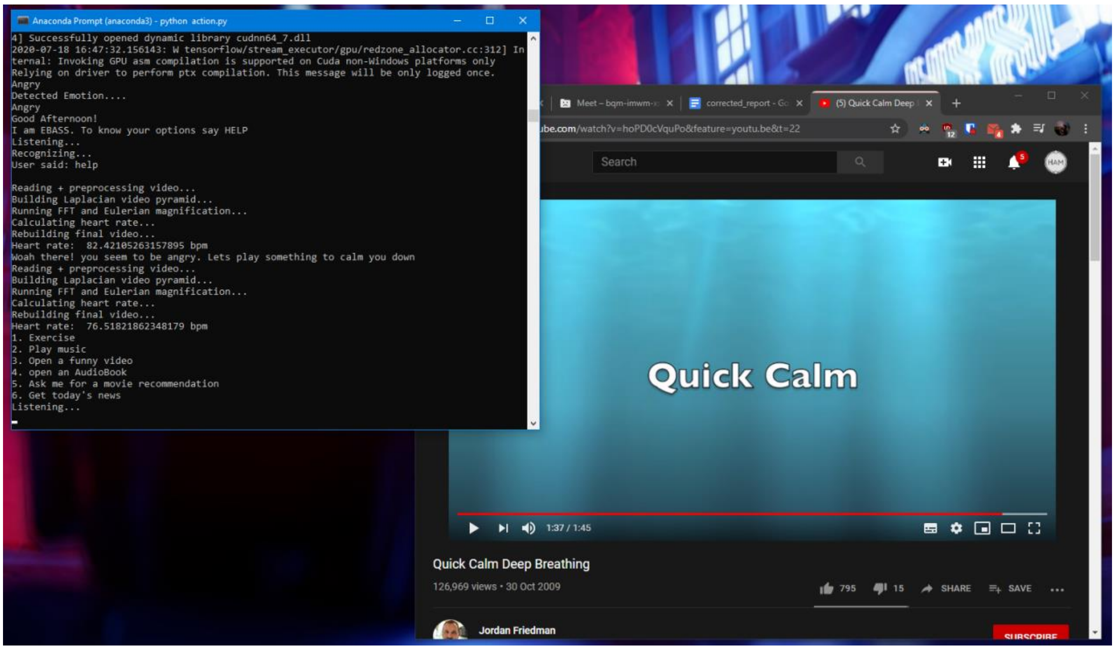
If the user is detected to be “Angry”, we start the Heart Rate Estimation module where we check the heart rate prior to our “Quick Calm” exercise.



```
Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  82.42105263157895 bpm
Woah there! you seem to be angry. Lets play something to calm you down
```

Figure 5.7: Heart rate estimation for an active video.

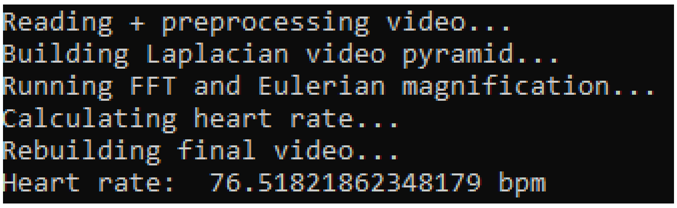
Here as shown in Fig 5.7, heart rate estimation was performed on an active video and it was found to be 82 bpm. Then, we redirect the user to a “Quick Calm” exercise.


Anaconda Prompt (anaconda3) - python action.py
Meet – bqm-imwm-
corrected_report - G
(5) Quick Calm Deep
ube.com/watch?v=hoPD0cVquPo&feature=youtu.be&t=22
Search
Quick Calm
1:37 / 1:45
Quick Calm Deep Breathing
126,969 views • 30 Oct 2009
795
15
SHARE
SAVE
Jordan Friedman
SUBSCRIBE

```
4] Successfully opened dynamic library cudnn64_7.dll
2020-07-18 16:47:32.156143: W tensorflow/stream_executor/gpu/redzone_allocator.cc:312] In
ternal: Invoking GPU asm compilation is supported on Cuda non-Windows platforms only
Relying on driver to perform ptx compilation. This message will be only logged once.
Angry
Detected Emotion....
Angry
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  82.42105263157895 bpm
Woah there! you seem to be angry. Lets play something to calm you down
Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  76.51821862348179 bpm
1. Exercise
2. Play music
3. Open a funny video
4. open an AudioBook
5. Ask me for a movie recommendation
6. Get today's news
Listening...
```

Figure 5.8: Quick calm exercise.

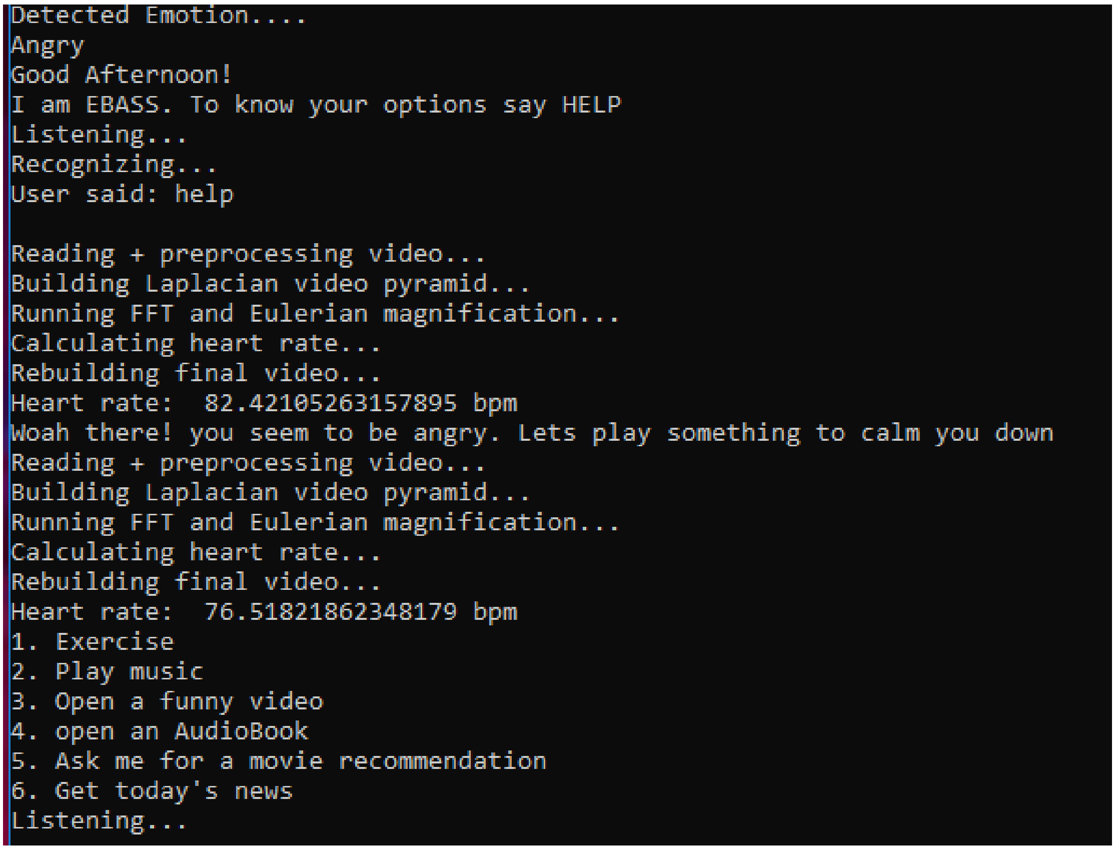
After the quick calm video has been played back, the user's heart rate is checked again and after a quick comparison with the previous heart rate check, one can understand that the heart rate has dropped significantly. Thereby, making the user calm and relaxed.



```
Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  76.51821862348179 bpm
```

Figure 5.9: Heart rate estimation after Quick calm exercise.

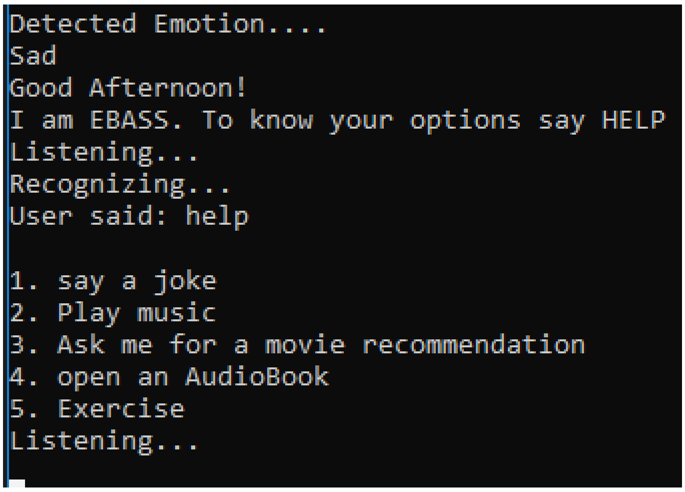
After running the Heart Rate Estimation again after the "Quick Calm" exercise, the heart rate was found to be 76 bpm.



```
Detected Emotion....
Angry
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  82.42105263157895 bpm
Woah there! you seem to be angry. Lets play something to calm you down
Reading + preprocessing video...
Building Laplacian video pyramid...
Running FFT and Eulerian magnification...
Calculating heart rate...
Rebuilding final video...
Heart rate:  76.51821862348179 bpm
1. Exercise
2. Play music
3. Open a funny video
4. open an AudioBook
5. Ask me for a movie recommendation
6. Get today's news
Listening...
```

Figure 5.10: Activity suggestion for Angry emotion

Figure 5.10 the list of activities suggested to an Angry person.



```
Detected Emotion....
Sad
Good Afternoon!
I am EBASS. To know your options say HELP
Listening...
Recognizing...
User said: help

1. say a joke
2. Play music
3. Ask me for a movie recommendation
4. open an AudioBook
5. Exercise
Listening...
```

Figure 5.11 : Activity suggestion for Sad emotion

Figure 5.11 shows the list of activities suggested to a Sad person

Chapter 6 - Conclusion and Future scope

6.1 Conclusion

Mental health is a major concern worldwide with rising awareness, it can be expected that early recognition and access to treatment will follow, as will the adoption of preventive measures.

Progress in mental health service delivery has been slow in most low- and middle-income countries and creation of a proper Emotion-based Assistant was something that was never done as per our survey. So, we came up with this project to tackle the mental health issue from its core. i.e- the emotion of a person.

The project has therefore been implemented only with minor issues such as our inability to compare our heart rate bpm results to an actual Pulse oximeter, Our database being limited to a very few number of activities as this is an initial release and can be further expanded as research is done upon the various mental issues that can arise. Our system also requires an expressive face to be fully functional, therefore a person with a non expressive face can find our system to be inefficient.

6.2 Future Scope

This project can act as a curtain raiser for others to see outside in a sense that its solid foundation leaves a room for plenty of further developments in improving the system so as to make it serve the people better.

- The project concerns with the detection of emotion and getting activities based off of the detected emotion,
- The whole system can be designed for a single particular user and based on his likings and preferences, the activities can be prioritized. This even includes the system “learning” about the user’s dislikes and similar disliked activities can be neglected.
- The system can also be able to classify the emotion based on the detected heartbeat of the user based on the factors that the heart rate depends upon such as Body to Mass Index (BMI), Age, Gender, etc.

- There is also the option to add a module such that a live video feed of the room or house can be shown to the user where their pet or child is displayed on screen. This may prove to be a real mood lifting activity.
- The system could be made to keep track of the past emotions and also track the activities that had helped the user to enhance his mood
- Inclusion of a timer, To-Do List or smart services such as weather reports can make this system feel closer to being a virtual assistant.
- Detect the emotion of the user while it runs in background to keep track of the user’s emotion while he/she performs their daily activities.

References

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479084/

[2] https://www.nimh.nih.gov/health/publications/men-and-depression/index.shtml

[3] https://www.helpguide.org/articles/relationships-communication/anger-management.htm

[4] https://www.mhanational.org/helpful-vs-harmful-ways-manage-emotions

[5] Tina Smilkstein et al. – “Heart Rate Monitoring Using Kinect and Color Amplification” - DOI: 10.1109/HIC.2014.7038874 2014

[6] Toshihiro Kitajima – “Heart Rate Estimation based on Camera Image”. Samsung R&D Institute Japan DT Lab Osaka, Japan - 2014 - DOI: 10.1109/ISDA.2014.7066275

[7] Wout Swinkels et al. – “SVM Point-based Real-time Emotion Detection”, Luc Claesen Faculty of Engineering Technology University Hasselt Diepenbeek, Belgium. 2017 - DOI: 10.1109/DESEC.2017.8073838

[8] Yubo Wang et al. – “Real Time Facial Expression Recognition with Adaboost”, - Dept. of Computer Science and Technology, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems ˈ Beijing 100084, PR China - 2004 - DOI: 10.1109/ICPR.2004.1334680

[9] Yingjie Chen – “A Comparison of Methods of Facial Expression Recognition”, Proceedings of the 1st WRC Symposium on Advanced Robotics and Automation Beijing, China 2018 - DOI: 10.1109/WRC-SARA.2018.8584202

[10] Karthik Subramanian Nathan et al. “EMOSIC - An Emotion Based Music Player For Android”, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, 2017 - DOI: 10.1109/ISSPIT.2017.8388671

[11] Salomi S. Thomas et al. - "Sensing Heart beat and Body Temperature Digitally using Arduino, Department of Computer Science and Engineering Dronacharya college of Engineering, 2016 - DOI: 10.1109/SCOPES.2016.7955737

[12] D. Yanga et al. "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment" Department of Computer Applications, National Institute of Technology, Haryana, India, 2017 - DOI: 10.1016/J.PROCS.2017.12.003

[13] Shlok Gilda et al. - "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation",Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, 2018 - DOI: 10.1109/WiSPNET.2017.8299738

[14] Prof. S.V. Kedar et al. - "Automatic Emotion Recognition through Handwriting Analysis: A Review", JSPM's RSCOE, S.P. Pune University, Pune, India, 2015 - DOI: 10.1109/ICCUBEA.2015.162

[15] Sofianita Mutalib et al. - "Towards Emotional Control Recognition through Handwriting Using Fuzzy Inference", SIG Intelligent Systems, Faculty of Information Technology and Quantitative Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia, 2008 - DOI: 10.1109/ITSIM.2008.4631735

[16] Seunghyun Yoon et al. - "MULTIMODAL SPEECH EMOTION RECOGNITION USING AUDIO AND TEXT", Dept. of Electrical and Computer Engineering, Seoul National University, Seoul, Korea, 2018 - DOI: 10.1109/SLT.2018.8639583

[17] Peixiang Zhong et al."EEG-Based Emotion Recognition Using Regularized Graph Neural Networks", 2020 - DOI: 10.1109/TAFFC.2020.2994159

[18] Esther Ramdinmawii et al. "Emotion Recognition from Speech Signal", Indian Institute of Information Technology Chittoor, Sri City, Andhra Pradesh, India, 2017 - DOI: 10.1109/TENCON.2017.8228105

[19] Hao-Yu Wu et al. - "Eulerian Video Magnification for Revealing Subtle Changes in the World" MIT CSAIL Quanta Research Cambridge, Inc.

[20] http://www.study-body-language.com/Face-expression-2.html

[21] Rapid Object Detection using a Boosted Cascade of Simple Features Paul Viola Michael Jones viola@merl.com mjones@crl.dec.com Mitsubishi Electric Research Labs Compaq CRL 201 Broadway, 8th FL One Cambridge Center Cambridge, MA 02139 Cambridge, MA 02142

[22] Arushi Raghuvanshi - “Facial Expression Recognition with Convolutional Neural Networks” Stanford University 2020 - DOI: 10.1109/CCWC47524.2020.9031283

[23] https://neurohive.io/en/popular-networks/vgg16/

[24] https://mlfromscratch.com/activation-functions-explained/#/

[25] https://developers.google.com/machine-learning/recommendation/content-based/basics