International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 5 ǁ 2021 ǁ PP. 45-55

Emotion Recognition Using Speech and Facial Expressions

Abstract:
Facial expression and voice have made significant progress in recent years with many commercial systems available for real-world applications. It gains strong interest to implement an emotion recognition system on a portable device such as tablet and smart phone device using the camera and microphone already integrated in the devices. It is very common to see face recognition phone unlocking app and voice in Google search engine in new smart phones which are proven to be hassle free way to unlock and search easily on a phone. Implementation of a emotion recognition system in a smart phone would provide fun applications that can be used to measure the mood of the user in their daily life or as a tool for their daily monitoring of the motion in phycology studies. However, traditional facial expression algorithms are normally computing extensive and can only be implemented offline at a computer. In this project, an automatic system has been proposed to recognize emotions from face images and speech on a smart phone in real-time. In our system, the camera of the smart phone is used to capture the face image and the microphone of the smart phone to capture voice, Personal Audio Classifier is used for the classification of voice features and Personal Image Classifier is used for the classification of Image features. The experimental results demonstrate that the proposed speech and facial expression recognition on mobile phone is successful and it gives up to 89.5% recognition accuracy.
Keywords: *Emotion Detection, Face Recognition, Audio Recognition, Convolutional Neural Network (CNN), Encoding, MIT Application*

Date of Submission: 27-04-2021 Date of acceptance: 11-05-2021

I. INTRODUCTION:

Face recognition technology has a wide range of applicability from auto unlock in smart phones and psychological usage in the past few years. It is a very good example on the success of computer vision on embedded devices. Facial expression recognition in such is a similar idea that will have numerous uses in the future. Facial expressions are one of those things that hold great importance to human in communication as they tend to convey emotions, energies and expressions without the use of words in communication. Darwin stated to the biological roots of facial expressions and the important role it play in the survival of the species including human beings. Ekman defined 5 basic emotional facial expression, happy, sad, anger, surprise and neutral and states how these expressions are universal. The deepest emotions are embedded in the facial expressions of human beings. Ekman states that all the expressions are a blend of all these basic expressions which sometimes makes it difficult to understand. Facial expressions basically function with the activation/ dilatation of one or more of the total forty-three facial muscles. The recognition of emotions by a smart phone, more precisely with the recognition of emotions from the speech which may be pitch, loudness, but also the spectral distribution of frequencies, for instance. Examples for applications where this is useful include call centres, or learning and game software. Knowledge about the emotional state can help to connect angry callers of an automatic dialogue system to a human operator, to motivate a student at the right t time, or to develop a fun game that is influenced by emotional expressions. However, emotion recognition from speech is a very challenging tasks. Other factors are the complexity of emotions as they may occur blended, or social influences may cause persons to shade or suppress their real emotional state. In order to recognize vocally expressed emotions automatically, affective information has to be separated from other influences on the voice, such as anomalies of the voice organs or physical effort which may be for instance a reason for breathlessness. For these reasons, recognition accuracies of current systems are still relatively low, so that affect recognition is hardly used in commercial products. Furthermore, it has not been possible to recognize arbitrary affect categories in real-time, which is necessary for most applications.

Existing System:

Emotion recognition by speech:

Several approaches to recognize emotions from speech have been reported. Most researchers have used global suprasegmental/prosodic features as their acoustic cues for emotion recognition, in which utterance- level statistics are calculated. For example, mean, standard deviation, maximum, and minimum of pitch contour and energy in the utterances are widely used features in this regard. Dellaert et al. attempted to classify 4 human

emotions by the use of pitch-related features. They implemented three different classifiers: Maximum Likelihood Bayes classifier (MLB), Kernel Regression (KR), and K-nearest Neighbors (KNN). Roy and Pentland classified emotions using a Fisher linear classifier. Using short-spoken sentences, they recognized two kinds of emotions: approval or disapproval. They conducted several experiments with features extracted from measures of pitch and energy, obtaining an accuracy ranging from 65% to 88%. The main limitation of those global-level acoustic features is that they cannot describe the dynamic variation along an utterance. To address this, for example, dynamic variation in emotion in speech can be traced in spectral changes at a local segmental level, using short-term spectral features. In, 13 Mel- frequency cepstral coefficients (MFCC) were used to train a Hidden Markov Model (HMM) to recognize four emotions. Nwe et al. used 12 Mel-based speech signal power coefficients to train a Discrete Hidden Markov Model to classify the six archetypal emotions. The average accuracy in both approaches was between 70 and 75%. Finally, other approaches have used language and discourse information, exploring the fact that some words are highly correlated with specific emotions. In this project, model information is used as acoustic features as well as the duration of voiced and unvoiced segments.

Emotion recognition by facial expressions:

Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human affective states. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. Mase proposed an emotion recognition system that uses the major directions of specific facial muscles. With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise. Yacoob et al. proposed a similar method. Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per-frame, mid-level representation.

They classified the six basic emotions by the used of a rule-based system with 88% of accuracy. Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows. They also built a mid- and high-level representation of facial actions by using a similar approach employed in, with 89% of accuracy. Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978, using permanent and transient facial features such as lip, nasolabial furrow and wrinkles. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups. They modeled the face by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved. In this project, the extraction of facial features is done by the use of model. Therefore, face detection and tracking algorithms are not needed.

II. LITERATURE SURVEY:

In paper [1], Facial expression has made significant progress in recent years with many commercial systems are available for real-world applications. It gains strong interest to implement a facial expression system on a portable device such as smart phone device. This can be done with the camera integrated already in the smart phone. It is very common to see face recognition phone unlocking app in new smart phones which are proven to be hassle free way to unlock a phone. Implementation a facial expression system in a smart phone would provide fun applications. That can be used to measure the mood of the user in their daily life. However, traditional facial expression algorithms are normally computing extensive and can only be implemented offline at a computer. In this paper, automatic system has been proposed to recognize emotions from face images on a smart phone in real-time. In our system, the camera of the smart phone is used to capture the face image, Personal Audio Classifier is used for the classification of voice features and Personal Image Classifier is used for the classification of Image features. Thus using this facial expression we will be able to recognize the users emotion. The experimental results demonstrate that the proposed facial expression recognition on mobile phone is successful and it gives recognition accuracy

In paper [2] Secure Cloud Storage and File Sharing 1 Bharat S. Rawal and 2 S. Sree Vivek1 Department of Information Sciences and Technology Pennsylvania State University, Abington, PA 19001, USA 2 ICU Medical, Chennai, India, proposed a secure file sharing mechanism for the cloud with the disintegration protocol (DIP). The paper also introduces new contributions of seamless file sharing technique among different clouds without sharing an encryption key.

In paper [3] FACE DETECTION AND RECOGNITION USING OPENCV Mrs. Madhuram M, B. Prithvi

Kumar, Lakshman Sridhar, Nishant Prem, Venkatesh Prasad, Assistant Professor, Department of

Computer Science, SRM Institute of Science and Technology, Ramapuram, Chennai, India: developed a face detection and recognition system using python along with OpenCV package. This system contains free modules which are detection, training and recognition. Basically, the detection module detects the face which gets into the field of vision of the camera and saves the face in the form of an image in JPG format. Then the training module trains the system using Haar cascade algorithm.

In paper [4] A Survey paper for Face Recognition Technologies Ms. Manjeet Kaur, M. Tech. CSE, Assistant Professor RIEM, Rohtak discussed their study about human behaviour and features. How we can recognize a face with the help of computers is given in this paper. Also, different ways which are available for face recognition and what are the problems with each technology is discussed.

In paper [5] presents the study based on the automated visitor tracking management system. Their visitor management system was useful at those places where a large number of visitors come and visit like colleges, tourist places etc. Their Visitor management solutions provide an ID to visitors in soft copy format. All of the records of those visitors were stored in the database at the time of check-in. Their modern visitor management system was used for restricting the visitors from prohibited areas by sounding an alarm or through some notification or through S.M.S. at the time of their visit.

Proposed System:

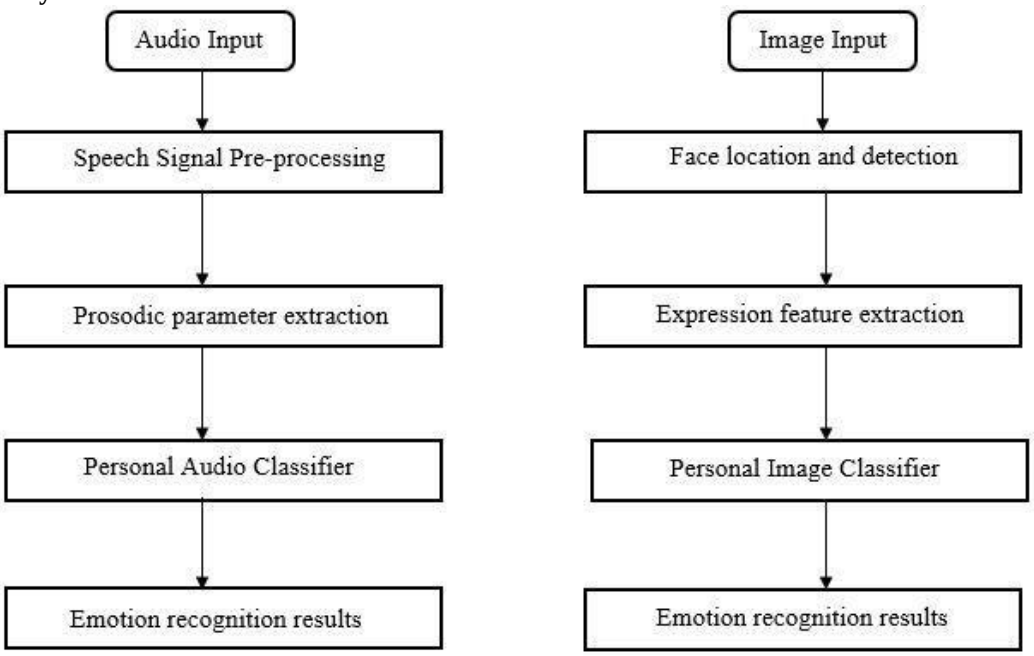

Fig 1. Proposed System

The proposed system deals with the emotion recognition by speech and facial expressions

- Stage 1: Image input and audio input is taken initially.
- Stage 2: Initially the input image is detected and speech signal pre-processing for the audio take place. The input image or the audio are moved for training. In this training phase, with the help of dataset, input image and audio are checked to match with the previously traineddataset.
- Stage 3: The input images are then classified with the help of classifier into different trained models according to their preferences.
- Stage 4: Extraction of exact output takes place and Personal image classifier and personal audio classifier classify the exact output of the image or audio.
- Stage 5: In this way the whole model works to detect the emotion through facial expression as well as speech. The output for the same is observed.

System Algorithm:

Using Personal Image Classifier for Face Emotion Recognition:

Training Stage:

1. In the first step, we make different sections and labels such as happy, sad, surprise, angry, neutral.
2. After labelling in each section add images related to the emotions labelled on it.
3. After adding images train the model.Testing Stage:

1. Each section in the model is tested to check we will give input as images with different emotions
2. To check each emotion, we will select each section for example to check happy give input as image showing happy emotion.
3. After giving the input click on test, it will show different emotion labels and the image which is related to that label it will highlight it as green or else red if not.

The distinction between the tested image and the image in the database is measured depending on the characteristic. Once the closest map to face is found, the image is recognized with a particular characteristic and the equivalent output is flashed on the screen.

Proposed Methodology:

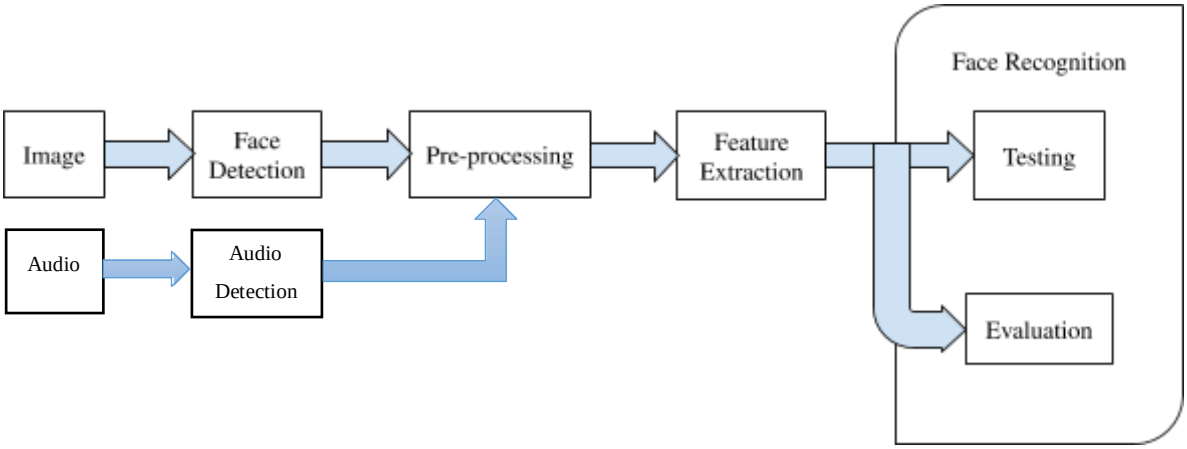

Fig 2. Phases of Face Recognition

1. Image Acquisition:

Image acquisition process initiates with capturing an image dynamically during runtime through the phone camera and in this procedure, it will store the images in the model. The captured image will be mapped with the existing image stored in the model and provide the required face image. The images will be stored in another directory and then comparison is made between captured images and images already stored in the directory.

2. Feature Extraction:

An extracted image has many landmarks, these characteristics provide description of the image. Personal Image Classifier generates characteristics of an image by taking a picture and transforming it into a collection of local feature vectors. With the help of model, characteristic vectors never change to any of the transformations of the image.

3. Face Emotion Recognition

After detection the capture image will be compared with the image stored in the model and it will give the required output.

4. Audio Emotion Recognition

After detection the capture audio will be compared with the audio stored in the model and it will give the required output.

Testing:

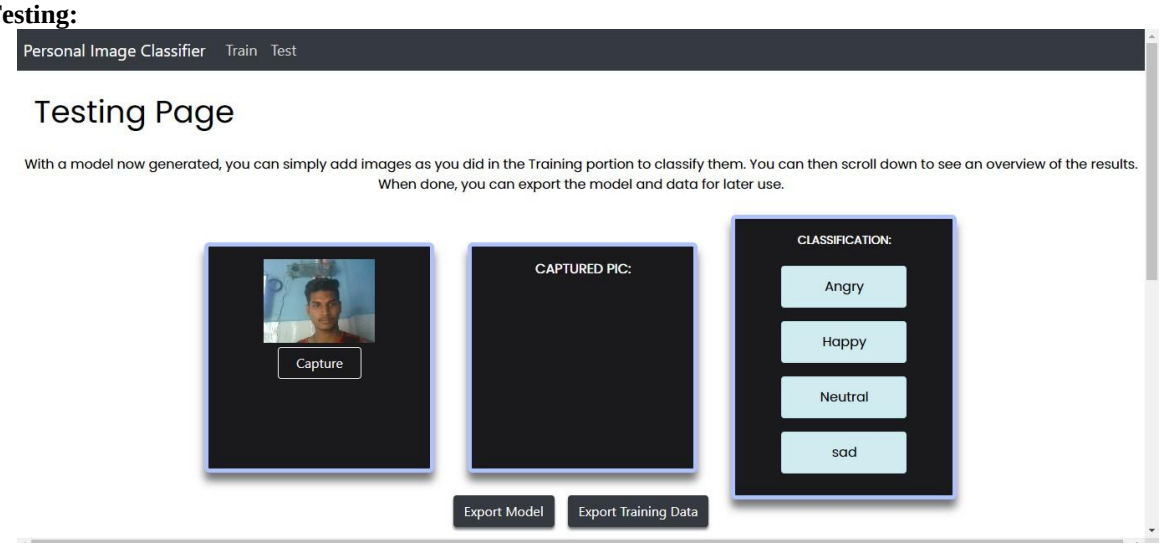

Fig 3. Testing Page

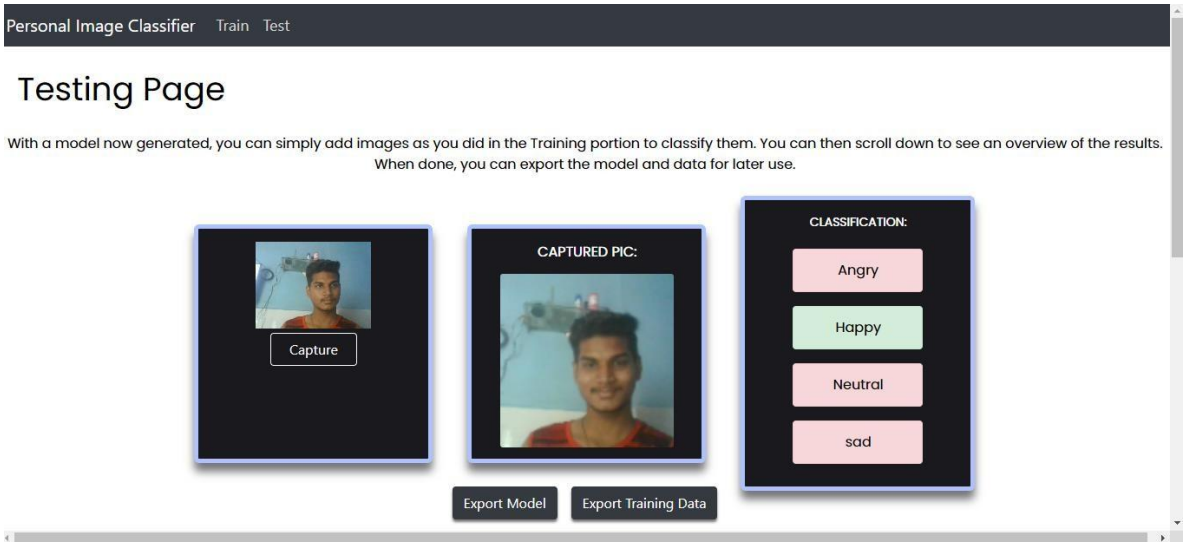

Fig 3.1. Image Testing Result

- To test an image is captured manually and the image is compared with the other images in the model. It will check the image with each image under each label and the image which matches the most it will display it with green colour or else in red.

Screenshots of the System:

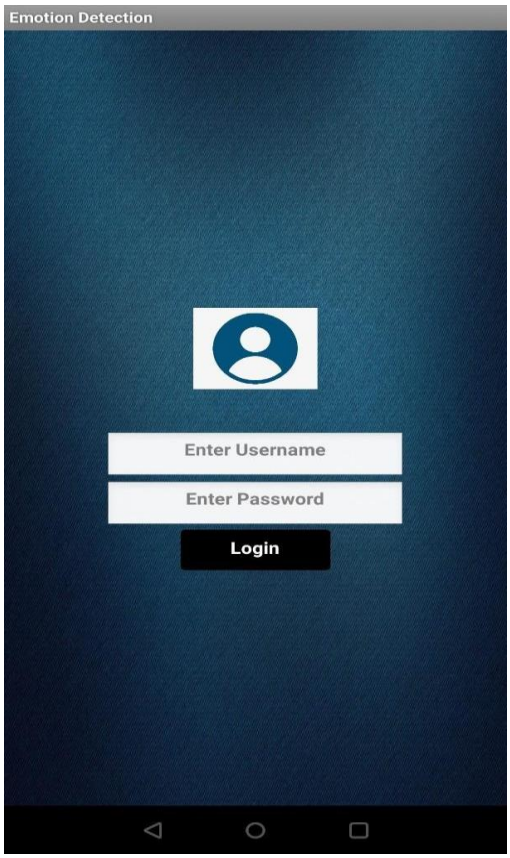

Fig 4. Login Activity

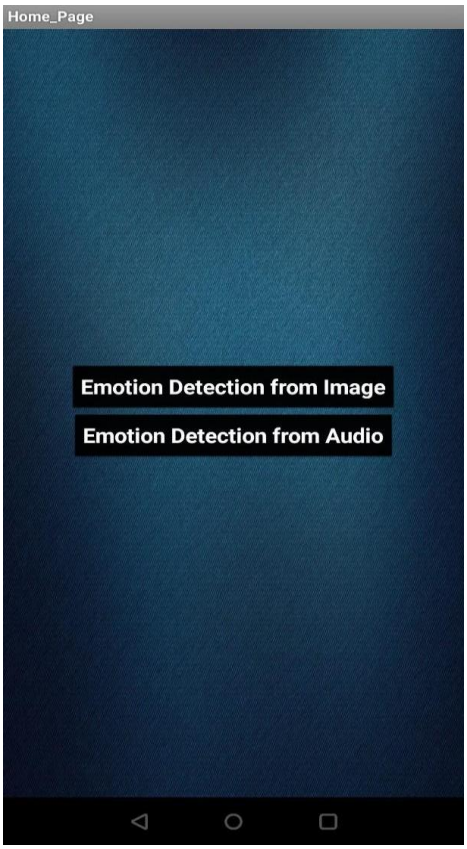

Fig 5. Menu List

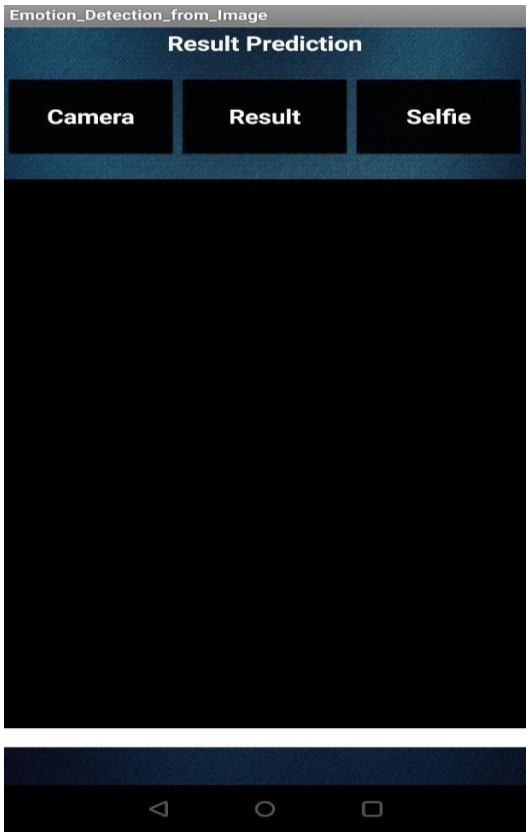

Fig 6. Main Screen for Emotion recognition through facial expression

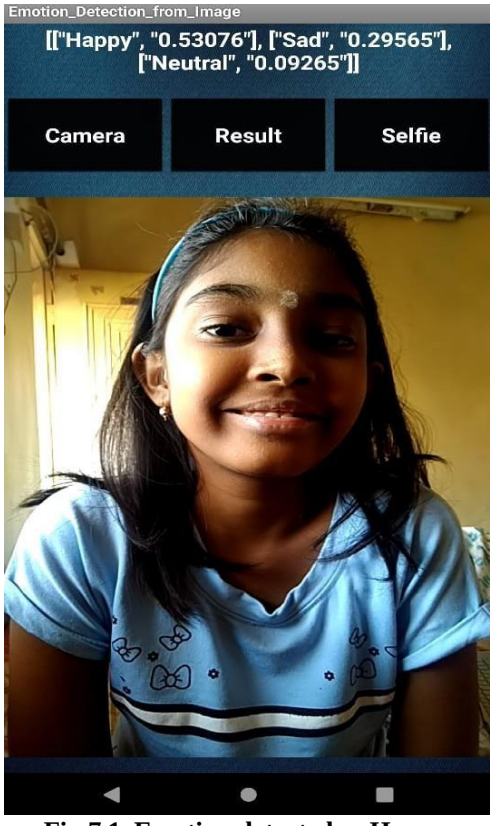

Fig 7.1. Emotion detected as Happy

Fig 7.2. Emotion detected as Neutral

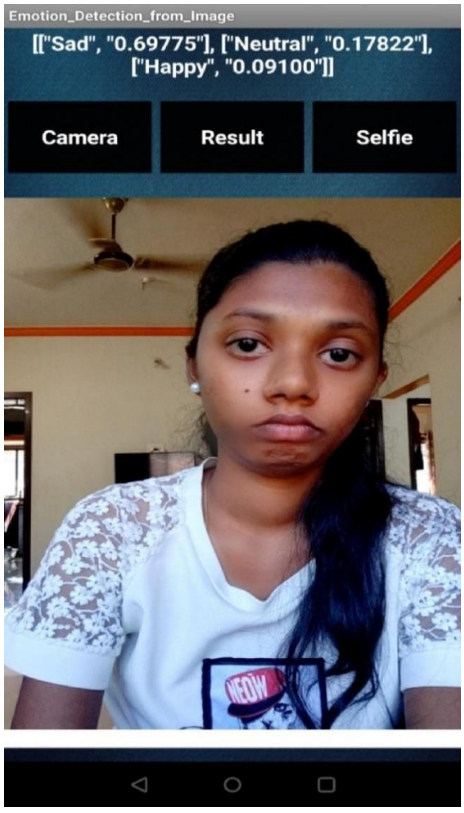

Fig 8. Emotion detected as Sad

Fig 9. Emotion detected as Surprise

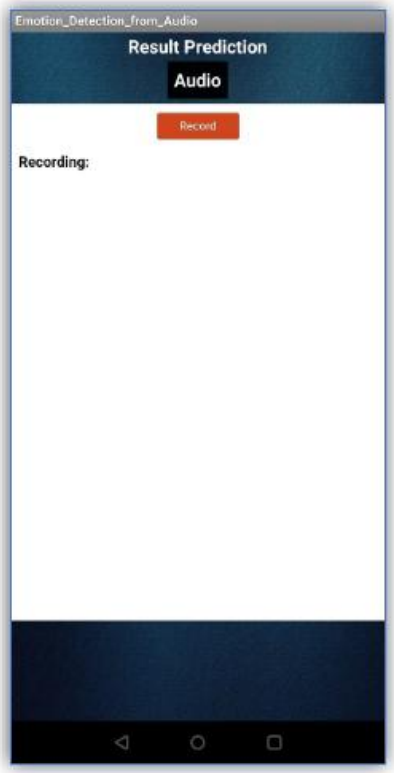

Fig 10. Main screen for Emotion recognition through Speech

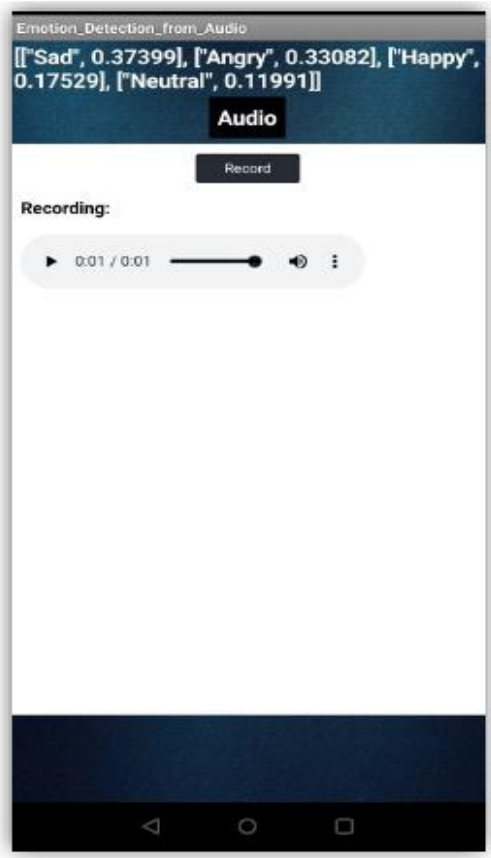

Fig 11. Emotion detected as Sad

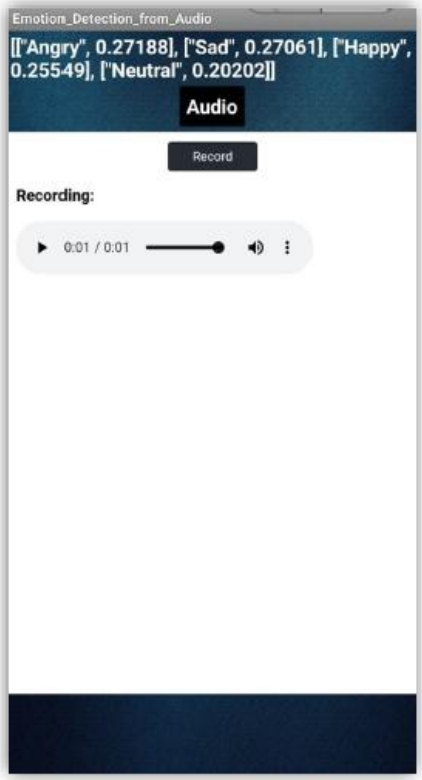

Fig 12. Emotion detected as Angry

III. CONCLUSION:

The system is built for emotion recognition from speech and facial expressions. The proposed approach tried to distinguish between six emotions (happiness, angry, sadness, surprise and neutral state) by using different classifiers. The results reveal that the system based on facial expression gave better performance than the one based on speech information only for the considered emotions. It can be observed that even though the system based on audio information had poorer performance than the facial expression emotion classifier, its features have valuable information about emotions, that cannot be extracted from the visual information. Audio and visual data present complementary information When these two modalities are used, the performance and the robustness of the emotion recognition system are improved. Further, the fusion performed at the feature level showed better results than the one performed at the score level. Gestures are widely believed to play an important role as well.