Facial Feature Analysis using Deep Convolutional Neural Networks and HAAR Classifier for Real-Time Emotion Detection

Bharath Surendar I
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

P Sri Siddharth Chakravarthy
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

Swarnalatha P
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

Roshini Thangavel
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

***Abstract*— Humans have several facial expression and all are dynamic in nature. In this paper, we have developed an application that comprises of an interface to perform facial emotion recognition in real-time. The model used in the application is developed using Deep Convolutional Neural Networks (DCNN), system It is designed to identify the emotion of the system user and then based on the emotion of the user the system will take precautions, if necessary, which involves sharing the geo-location of the user in case the system predicts the emotion of the user to be perilous. It will also help user to manage their emotions by podcasting some contents.**

Keywords*— Human Computer Interaction · Emotion recognition · Deep convolutional neural network · HAAR classifier · Mobilenet classifier · Deep learning · Feature extraction

I. INTRODUCTION

Deep Neural Networks (DNNs) outperform traditional models in numerous optical recognition missions containing Emotion Recognition which is an imperative process in next-generation Human-Computer Interaction (HCI) for behavioural description. Existing facial emotion recognition methods lag in terms of high accuracy and are not sufficient and not practical in real-time applications. This work proposes the use of Deep Convolutional Neural Network (DCNN) method for Facial emotion recognition in Images. The proposed network architecture consists of Convolution layers by which the model extracts the HAAR features from the input facial image and by using the deep convolutional neural network the temporal dependencies which exist in the images can be considered during the prediction.

Humans use facial expressions to point out their emotional states. However, countenance recognition has remained a challenging and interesting problem in computer vision. Extensive possibilities of applications have made emotion recognition ineluctable and challenging within the field of computing . the utilization of non-verbal cues like gestures, body movement, and facial expressions convey the sensation and therefore the feedback to the user. This discipline of Human–Computer Interaction places reliance on the algorithmic robustness and therefore the sensitivity of the sensor to ameliorate the popularity . Sensors play a big role in accurate detection by providing a really high-quality input, hence increasing the efficiency and therefore the reliability of the system. Automatic recognition of human emotions would help in teaching social intelligence within the machines. This paper presents a quick study of the varied approaches and therefore the techniques of emotion recognition. The survey covers a succinct review of the databases that are considered as data sets for algorithms detecting the emotions by facial expressions. Later, mixed reality device Microsoft HoloLens (MHL) is introduced for observing emotion recognition in Augmented Reality (AR). a quick introduction of its sensors, their application in emotion recognition and a few preliminary results of emotion recognition using MHL are presented.

As the years of intensive research and development elapsed, computers became rather more powerful than before. Recently, newer algorithms and techniques are being developed for image pre-processing, feature extraction, and powerful classification methods. Facial recognition has becomes a major feature for industries leading in mobile phones, security chains, etc.

The objective of the Deep Convolution Neural Network is to identify the visual input and predict the corresponding character. It is used to figure out an image and understand how pixel elements are arranged in an image. It could be holding the input image, deducting a weight matrix and the input convoluted to specify features from the image without considering about the dimensional arrangement and decreasing the number of specifications from the visual input. The convoluted images have lesser pixels when compare to an original input. It is definitely decreasing the number of parameters needed to train the network. Facial emotion recognition involves the process of interpreting the visual input and then print the corresponding predicted emotion. When it comes to visual real-time facial analysis the task is difficult to process due to the complexity in the image in face size, brightness of the room, etc. In the proposed system the scanned image is pre-processed and segmented into paragraphs, paragraphs into lines, lines into words and words into character image glyph. Then we perform feature extraction methods to extract features from these character image glyph to extract the features such as character height, width, number of horizontal and vertical line patterns, horizontally and vertically oriented curve patterns, circles, number of slope lines, image centroid and special scripts.

II. LITERATURE REVIEW

Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do. AI can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Emotional artificial intelligence, or emotion AI, is a technology that is capable of reading, imitating, interpreting, and responding to human facial expressions and emotions. Understanding contextual emotion has widespread consequences for society and business. In the public sphere, governmental organizations could make good use of the ability to detect emotions like guilt, fear, and uncertainty. It's not hard to imagine the TSA auto-scanning airline passengers for signs of terrorism, and in the process making the world a safer place.

Facial expression recognition(FER) deals with the process of identifying face expressions associated with human emotions that are displayed when a person exhibits an emotion. FER is believed to have several applications especially in the fields of medicine, human computer interaction, surveillance, etc. where recognition of a person's emotion based on the facial expressions is vital. The process of detecting emotions from face expression generally involve two fundamental steps mainly feature extraction and emotion recognition [1]. There have been research studies in the field of FER by using DCNN models as the base for developing the system. The research study that worked on the Child Affective Facial Expression (CAFE) dataset and the Extended Cohn Kanade (CK+) dataset by Adish Rao explored the possibility of developing an emotion recognition technique that works for both adults as well as children by identifying the differentiating features. The proposed model used an integrated DCNN architecture that incorporated facial landmarks to predict the emotions. This research focused on the estimation of the optimal facial landmarks that were needed in order to obtain desirable results for a FER system [2]. However the use of DCNN for FER systems comes with a bottleneck. DCNN is considered to be the optimal approach for most effective and efficient facial emotion recognition systems. However, DCNN based approach has its own downside. The use of DCNN for FER systems need to be properly implemented in order to overcome the bottleneck that arises due to stability and infinite feasibility problems. These problems arise when we deal with human faces of different races. Researchers have been able to overcome this issue by using a novel feature extraction method for such bottleneck conditions [3]. Apart from FER systems recently researchers have also began working on speech emotion recognition (SER) using DCNN. The drawbacks on SER systems is the time that needs to dedicated to annotate the data can be exorbitant, thus when the model which is trained on such limited data samples fails to perform on newer data samples. The study can also be associated to the case of FER where there is a possibility of getting wrong predictions, but by using active learning (AL) approach we can select active samples using greedy sampling and uncertainty methods to train the model and evaluate the performance [4].

Apart from traditional approaches to develop a FER system using DCNN models Ismail Shahin and set of researchers proposed a text and speaker independent emotion recognition system that used a novel classifier. The model proposed was a hybrid architecture that composed a hybrid cascaded gaussian mixture and deep neural network (GMM-DNN) model in a single architecture. The researchers used the Speech database (from the Arabic U.A.E Database) that consisted six unique emotions. The proposed architecture was tested against traditional emotion classifiers like SVMs (support vector machines) and MLPs (Multilayer Perceptron) classifiers and the results proved that the hybrid model outperformed the traditional classifiers with a perform accuracy of 83.97%. The same model when tested with databases with noisy conditions also performed significantly better [5]. Developing a FER system is hectic but to program an automated FER system can be challenging as well. Recent breakthrough in the field of computer vision have supported the improvements in FER systems drastically. Particularly the DCNN based approach has shown substantial changes and improved results, the reason that DCNN can outperform Convolutional Neural Network based facial analysis classifiers is due to two key factors a. the traditional CNN architecture involves tuning of parameters to tolerate mixture and complementarity in results, b. the classification rule embedded in FER developed using CNNs lack quality. Thus, when training the model on a posterior probabilities allow capture of non-linear dependencies among classification rules [6]. Using FER for surveillance involves recognizing facial emotions from video files. To extract facial features from video frames we tend to make use of the local characteristics (landmarks) to produce geometric-based facial features that help to discriminate between the various emotion expressions. A study conducted by Amr Mostafa involved the development of a FER system for the BioVid Emo database [7] using machine learning classification algorithms such as random forest (RF), recurrent neural network (RNN), etc.

Face recognition is of great importance to world applications like video surveillance, human machine interaction and security systems. As compared to traditional machine learning approaches, deep learning based methods have shown better performances in terms of accuracy and speed of processing in image recognition. The leading edge innovation of countenance Recognition framework is that the consumer satisfaction estimation. MFER, a completely unique procedure is proposed during this paper for identifying consumer satisfaction levels. The model must anticipate client's behaviour within the dynamic cycle. As stated before the traditional facial emotion recognition systems help only with the classification of basic emotions, and these fundamental emotions are limited to express only restrained and contrasting emotions.

Researchers have been working on newer methods to develop models that can understand and interpret complex facial emotions, one such approach is discussed by Yong Yang and fellow researchers by making use of valences or the concept of thresholds for each class of emotions [8]. They worked on an arousal-valence continuous emotion space model, to enhance existing emotion recognition systems. In their approach they have used arousal reflects that corresponds to the emotional intensity of the person, the main purpose of using the valences is to classify positive and negative emotions. This approach assigned values from -1 to 1 for each emotion based on its intensity and the trained the model using an integration of CNN for extracting the features from the facial data and using support vector regression(SVR) model to predict the emotion. The experimental results from their research showed that their approach exhibited better recognition result than the traditional methods.

Deep Convolutional Neural Network (CNN) may be a special sort of Neural Networks, which has shown exemplary performance on several competitions associated with Computer Vision and Image Processing. a number of the exciting application areas of CNN include Image Classification and Segmentation, Object Detection, Video Processing, tongue Processing, and Speech Recognition [9]. The powerful brain of deep CNN is primarily thanks to the utilization of multiple feature extraction stages which will automatically learn representations from the info . the supply of an outsized amount of knowledge and improvement within the hardware technology has accelerated the research in CNNs, and recently interesting deep CNN architectures are reported. Several inspiring ideas to bring advancements in CNNs are explored, like the utilization of various activation and loss functions, parameter optimization, regularization, and architectural innovations [10]. However, the many improvement within the representational capacity of the deep CNN is achieved through architectural innovations. Notably, the ideas of exploiting spatial and channel information, depth and width of architecture, and multi-path information science have gained substantial attention. Similarly, the thought of employing a block of layers as a structural unit is additionally gaining popularity.

III. PROPOSED WORK

The goal of this project is to capture the face of the user and predict the emotion of the user by analysing the facial expressions from the captured image. Human emotions are very difficult for computers to understand due to the fact that the overall expression for the same emotion differs from person to person based on the face. This difference could be due to the fact that every person has a different facestructureand some might not display the emotions prominently compared to others. In general, there are 7 types of human emotions that are recognised universally namely anger, happiness, disgust, fear, sad, surprise and neutral.

A. Preparing the dataset

First we need to prepare a dataset with which we train the model. The Model we will be using contains different datasets of faces for example CK+ dataset and some images. Once the dataset is fixed we then train the model by passing the train dataset through the convolutional layers. Once the model is trained we will test with model with test inputs which is the real-time feed from the user's webcam, the trained model will be able to recognize emotions from the input image.

B. Detecting faces

The system gets the user's face from the webcam feed which is then passed onto the model for predicting the emotion. The emotion from the user's face is predicted using HAAR cascade classifier, which is a course work in OpenCV that is used for object recognition. HAAR classifier are mainly used for face recognition. When the image is passed to the HAAR classifier the image is changed to grey -scale and is resized to a similar size as the images in dataset. A HAAR course is fundamentally a classifier which is used to detect specific objects from the source.

C. HAAR Classifier

By using HAAR cascade classifier we will be able detect the face from the image and then by using keras model's Deep Convolutional Neural Network we can classify the emotion recognized from the user's face.

D. Parsing the data and emotions

We will parse the data frame-by-frame and the output will be produced in the form of a text file that will contain emotions captured frame-by-frame from the input. Now, we will set a limit for certain emotion like fear for which it sends the data if the limit is crossed.

E. System Architecture

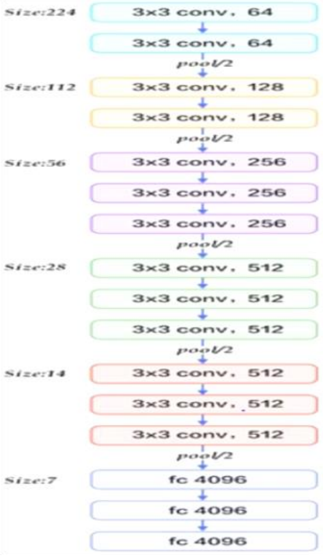


Fig. 1. DCNN model –VG16

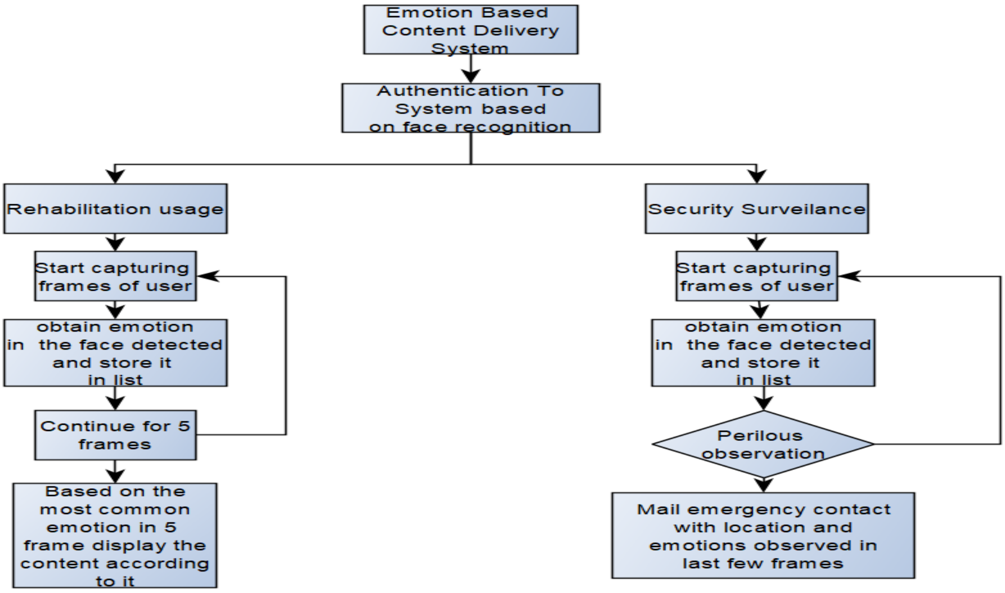
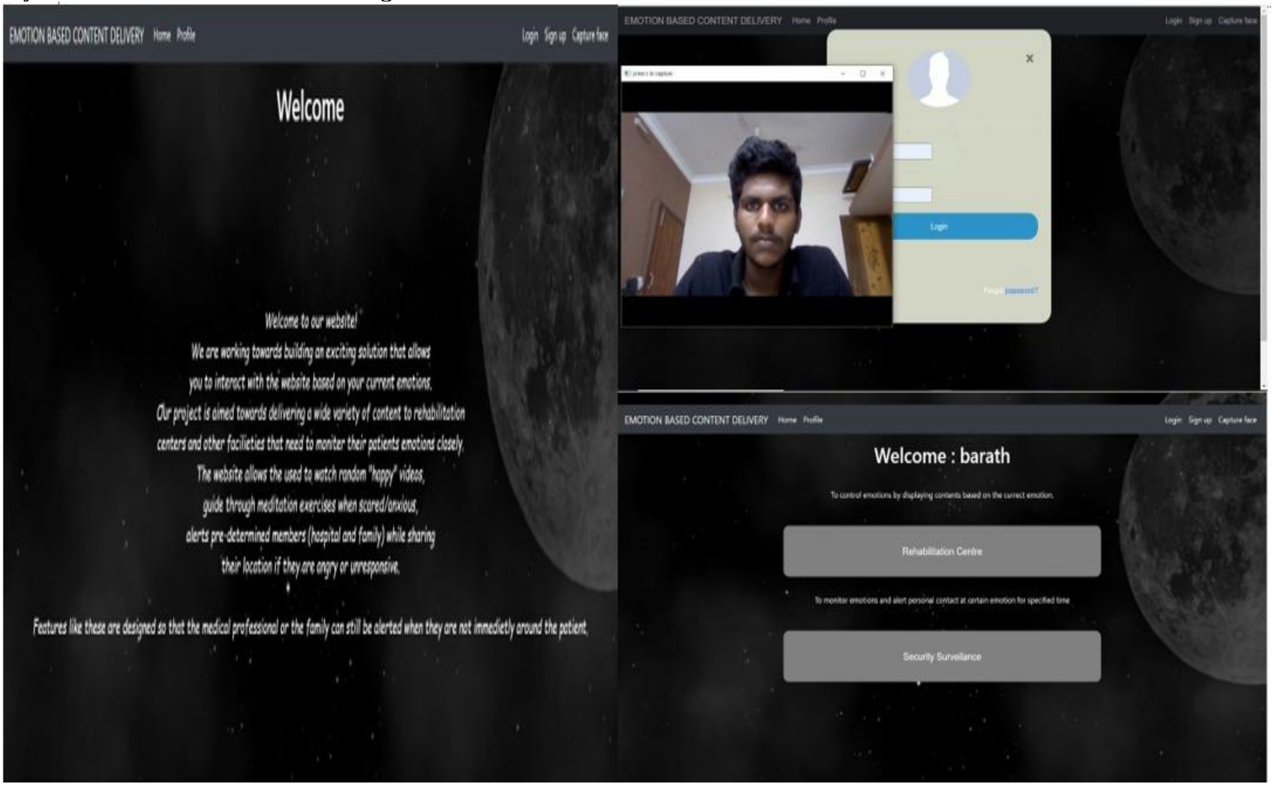

Emotion Based Content Delivery System
Authentication To System based on face recognition
Rehabilitation usage
Security Surveilance
Start capturing frames of user
obtain emotion in the face detected and store it in list
Continue for 5 frames
Based on the most common emotion in 5 frame display the content according to it
Start capturing frames of user
obtain emotion in the face detected and store it in list
Perilous observation
Mail emergency contact with location and emotions observed in last few frames

Fig. 2. User Interface System Architecture

IV. Results

After developing the classification model and the interface we have got the system working as expected and extracted its full application. As seen in Figure 3 the emotion detection program has been integrated into a website for easy use in rehabilitation centres and other facilities. This program can be accessed using the UI as shown in Figure 3 by the centres to log in and access the patient information. The patients input (their image) and output (response based on the emotion) will be visible as seen in Figure 4. The website has been designed especially for rehabilitation centres of drug and alcohol abuse victims who are recovering to monitor their moods and emotions and provide a rapid response whenever necessary. This system not only keeps track of all the emotions detected on the patient but also provides an apt response based on the emotion to help the patient in their recovery process. In the case of an emergency, where the patient is facing an adverse emotion, the system triggers an emergency response email with all the information necessary to the registered contact to be able to take the necessary measures. This response includes an email with the location of the patient, the map of the location, as well as the past emotions faced by the patient as seen in Figure 5.



Welcome to our website!
We are working towards building an exciting solution that allows
you to interact with the website based on your current emotions.
Our project is aimed towards delivering a wide variety of content to rehabilitation
centers and other facilieties that need to moniter their patients emotions closely.
The website allows the used to watch random "happy" videos,
guide through meditation exercises when scared/anxious,
alerts pre-determined members (hospital and family) while sharing
their location if they are angry or unresponsive.

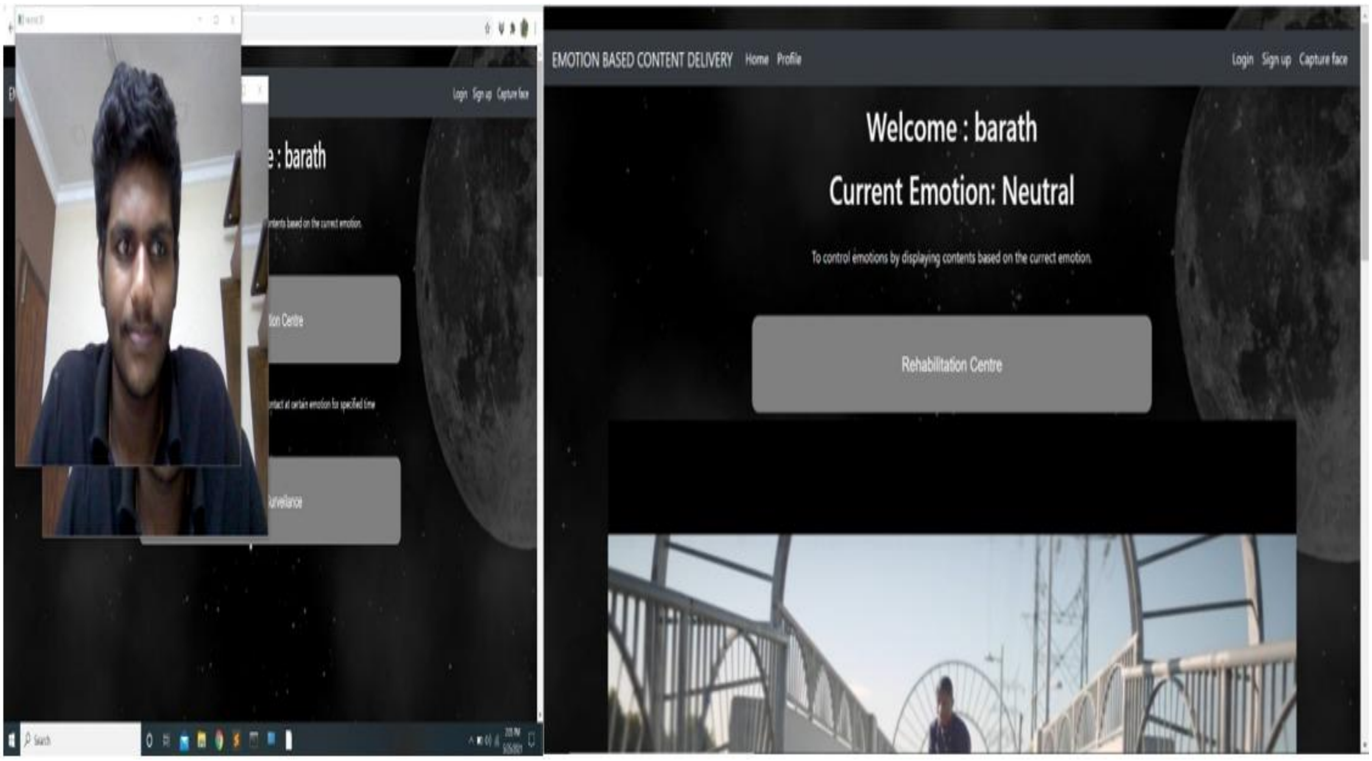
Features like these are designed so that the medical professional or the family can still be alerted when they are not immedietly around the patient.

Welcome : barath

To control emotions by displaying contents based on the currect emotion.

To monitor emotions and alert personal contact at certain emotion for specified time

Fig. 3. User Interface of Emotion Based Content Delivery Website



EMOTION BASED CONTENT DELIVERY Home Profile Login Sign up Capture face

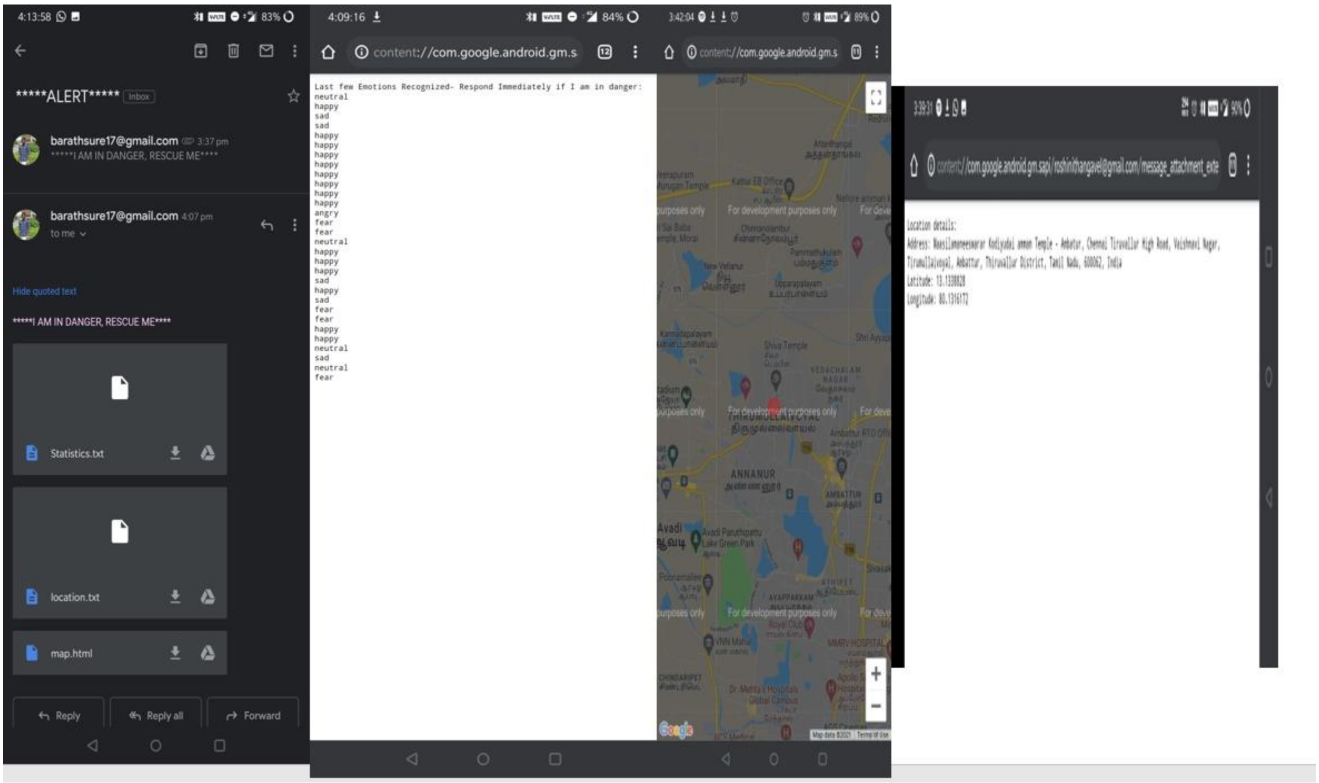
Welcome : barath

Current Emotion: Neutral

To control emotions by displaying contents based on the current emotion.

Rehabilitation Centre

Fig. 4. Input and Output UI of Rehabilitation centre application



*****ALERT***** Inbox

barathsure17@gmail.com 3:37 pm
*****I AM IN DANGER, RESCUE ME****

barathsure17@gmail.com 4:07 pm
to me

Hide quoted text

*****I AM IN DANGER, RESCUE ME****

Statistics.txt

location.txt

map.html

Reply Reply all Forward

content://com.google.android.gm.s

```
Last few Emotions Recognized- Respond Immediately if I am in danger:
neutral
happy
sad
sad
happy
happy
happy
happy
happy
happy
happy
happy
angry
fear
fear
neutral
happy
happy
happy
sad
happy
sad
fear
fear
happy
happy
neutral
sad
neutral
fear
```

content://com.google.android.gm.sapi/roshinithangavel@gmail.com/message_attachment_exte

```
Location details:
Address: Masilamaneeswarar Kodiyudai amman Temple - Ambatur, Chennai Tiruvallur High Road, Vaishnavi Nagar,
Tirumullaivoyal, Ambattur, Thiruvallur District, Tamil Nadu, 600062, India
Latitude: 13.1338828
Longitude: 80.1316172
```

Fig. 5. Mail Generated and the contents of it in Security Surveilance application

A. Comparisions between various models available

TABLE I. ACCURACY BETWEEN MODELS

MODEL	IMAGE NET ACCURACY	MILLION MULT ADDS	OTHER PARAMETERS (in millions)
VGG16	71.5%	15300	138
GOOGLENET	69.8%	1550	6.8
MOBILENET	70.6%	569	4.2

TABLE II. TENSORFLOW NETWORK COMPARISON(BASED ON CLASSIFICATION TIME)

NETWORK	CLASSIFICATION TIME (TENSOR FLOW 1.5)	CLASSIFICATION TIME (TENSOR FLOW 1.8)
DenseNet201	658	628
InceptionV3	175	184
ResNet	301	299
MobileNet	131	135

While training, the performance of the model can be viewed and analysed using TensorBoard in the local host which is shown in the below figure. The accuracy and other detailed statistics can also be analysed and visualized by using the TensorBoard.

B. Accuracy and Cross Entrophy

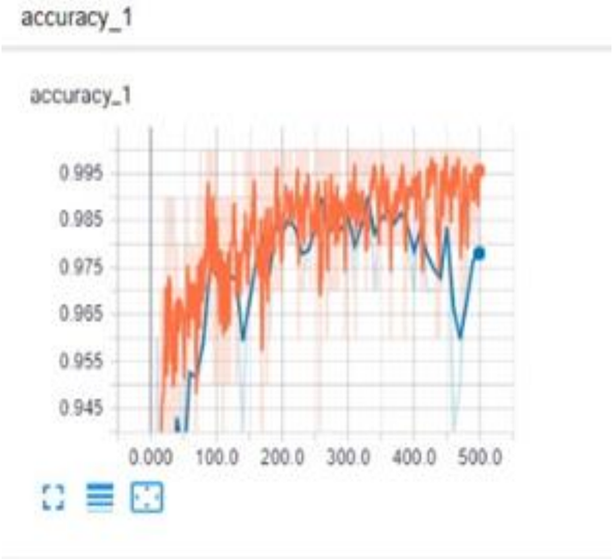


Fig. 6. Accuracy of the model

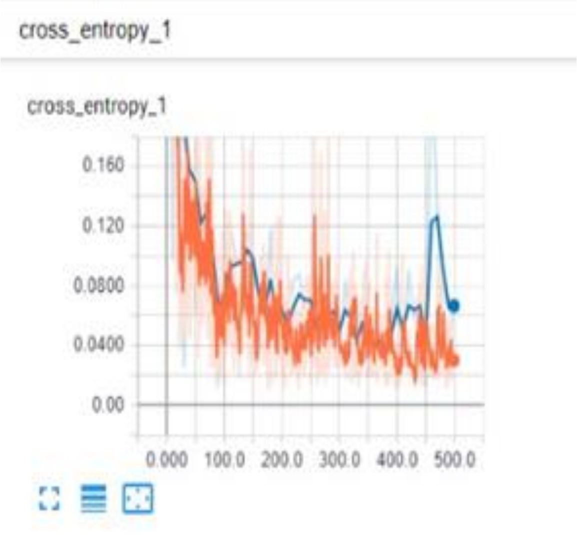


Fig. 7. Cross Entrophy of the model

The two graphs represent the accuracy and cross entropy. The accuracy graph depicts how well the model is classifying in the training phase. we can see that the accuracy gradually increases (orange line represents the accuracy for the training data and the blue line indicates the accuracy of the validation data). Whereas the second half of the system shows the cross entropy which represents the learning of the model the network weights gets updated as the training gets processed. Here the cost function is decreased. Cost function represents the difference between the actual and the predicted outcome. The Following graphs shown above are tracked down during the training process and in order to visualize these graphs we have used tensor board.

V. CONCLUSION

Human emotions can describe a lot about what a person is feeling at the moment. By this research we were able to develop a user-interface in form of a website, where a user can login and we can predict the emotion expressed by the user currently. The Interface gives two options, one based on the current emotion it podcasts some video to make the user happy all the time and some other videos to control their current emotion. Other is if the person is in perilous condition it alerts the user’s personal contact with location and emotions recognised for specific period of time.

VI. FUTURE SCOPE

The purpose of a Deep Convolutional Neural Network based facial emotion classifier is to apply the proposed system to real-time use. We believe that our model can be used in rehabilitation centres to monitor patients and to see that they do not do take any impulsive decisions. We can also extend the use of emotion detection to recommender systems such song suggestion or movie suggestion based on the mood the user is in. The main motive would be to help overcome the social trauma that the current generation of social media elements caused on the users.

REFERENCES

[1] Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. H. (2020, March). Facial emotion recognition using deep convolutional neural network. In *2020 6th International conference on advanced computing and communication Systems (ICACCS)* (pp. 317-320). IEEE.

[2] Rao, A., Ajri, S., Guragol, A., Suresh, R., & Tripathi, S. (2020). Emotion Recognition from Facial Expressions in Children and Adults Using Deep Neural Network. In *Intelligent Systems, Technologies and Applications* (pp. 43-51). Springer, Singapore

[3] Ma, T., Benon, K., Arnold, B., Yu, K., Yang, Y., Hua, Q., ... & Paul, A. K. (2020, November). Bottleneck Feature Extraction-Based Deep Neural Network Model for Facial Emotion Recognition. In *International Conference on Mobile Networks and Management* (pp. 30-46). Springer, Cham.

[4] Abdel Wahab, M., & Busso, C. (2019, September). Active learning for speech emotion recognition using deep neural network. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1-7). IEEE

[5] Shahin, I., Nassif, A. B., & Hamsa, S. (2019). Emotion recognition using hybrid Gaussian mixture model and deep neural network. *IEEE access*, *7*, 26777-26787

[6] G. Pons and D. Masip, "Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis," in *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 343-350, 2018

[7] A. Mostafa, M. I. Khalil and H. Abbas, "Emotion Recognition by Facial Features using Recurrent Neural Networks," *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, Egypt, 2018, pp. 417-422.

[8] Y. Yang and Y. Sun, "Facial Expression Recognition Based on Arousal-Valence Emotion Model and Deep Learning Method," 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), Dalian, China, 2017, pp. 59-62

[9] Ouyang, X., Kawaai, S., Goh, E. G. H., Shen, S., Ding, W., Ming, H., & Huang, D. Y. (2017, November). Audio-visual emotion recognition using deep transfer learning and multiple temporal models. In *Proceedings of the 19th ACM international conference on multimodal interaction* (pp. 577-582).

[10] Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 443-449). ACM.