Vellore Institute of Technology, Vellore

School of Computer Science and Engineering (SCOPE)

REPORT

FACIAL EMOTION THRESHOLD BASED DATA DISPATCHER

By:

Reg. No:	Name
18BCE0541	Kurakula Tanu Sri
18BCE0624	Kurakula Rupa Sri
18BCE0595	Aditya Singh
18BCE0606	Kshitiz Choudhary

To:

Prof. Shashank Mouli Satapathy

for the course entitled as

"HUMAN COMPUTER INTERACTION"

School of Computer Science & Engineering

NOVEMBER, 2020.

TABLE OF CONTENTS

1. INTRODUCTION....4
2. RELATED WORK....4
3. REAL LIFE APPLICATION....9
4. INDIVIDUAL CONTRIBUTION....9
5. TOOLS AND TECHNOLOGIES USED....9
6. PROPOSED SYSTEM PROCESS FLOW....10
7. WORKING METHODOLOGY....11
8. IMPLEMENTATION RESULTS AND USER INTERFACE....11
9. INTERFACE VALIDATION....16
10. COMPARITIVE ANALYSIS....17
11. CONCLUSION AND FUTURE SCOPE....18
12. REFERENCES....19
13. APPENDIX....20

ABSTRACT

Predicting the human emotions is not so easy without analysing their expressions on the face. Face emotions are one of the most informative gestures. So, we make use of these gestures on the face for the emotion recognition of the person. The state of the response can be decided based on the mood of the opposite person. To continuously monitor the expressions of a person and predicting the emotions of that person is difficult. So, the automation process makes it completely flexible to analyse and dispatch the emotions of person in the form of data. The project is about facial detection and emotional recognition of a person using Convolutional neural network (CNN). In facial recognition, the model recognizes the face of a person and in the emotion recognition the model recognizes the facial expression of that particular person using feature extraction and the nodal points. This is a hybrid model of facial emotional recognition and data dispatcher. It records the feelings of an individual in each frame of a video that is taken by means of webcam and dispatch the information. Dispatching of emotional data can be used for alerting the status of a person's emotion via a means of message (mail or message) during certain period of time. Convolutional neural network has outperformed in the field of image classification of emotion of a person as it is very dynamic in nature and the mood changes within a second. This project can analyse seven different facial expressions such as happy, sad, disgust, anger etc in real-time. The main objectives include classifying and predicting emotions in real time, storing the data of person indicating the status of the emotions and dispatch data when needed. This model tries to develop these decision-making and classification skills by training the machine. Neural network has been used to get the better results. This project mainly focuses on market research domain that can get the customer feedback. It can be used to find their interests that helps to improve the marketing strategy.

1. INTRODUCTION

It is important to know a person's emotion while they are doing something. It can be used to find their interest in doing it. Our brains have neural networks which are responsible for all kinds of thinking (decision making, understanding). This model tries to develop these decisions making and classification skills by training the machine. It classifies multiple faces and predict different emotions at the same time. In order to obtain higher accuracy, we take the models which are trained over thousands of datasets.

Charles Darwin, the first scientist stated that recognizing facial expression is one of the most powerful and immediate means for human beings to communicate their emotions, intentions and opinions to each other. Facial expression can also provide information about cognitive state, such as interest, boredom, confusion, and stress. Human life is a complex social structure. For humans it is not possible to navigate without reading the other persons. They do it by identifying the faces. The state of response can be decided based on the mood of the opposite person. A person's emotion can be figured by observing his emotion. This concept can be used in many areas such posting ads or what kind of products they are more interested, by continuously observing their expression in whatever they watch, if they are watching it with more happiness and surprised, then we can say that they are more interest in it.

Face Recognition and Facial emotion recognition are similar problems. In the face recognition system, the model recognizes using the face expression of the person Whereas in facial emotion recognition, the model identifies the emotion of the face expressed by someone using facial expression and also the nodal points. The recent trends show that deep learning applications have gotten noticeable performance in computer vision applications. and therefore, the convolutional neural networks have outperformed within the field of image classification.

2. RELATED WORK

Marian Stewart et al [1], have proposed a concept on a user-independent fully automatic system for real-time recognition of basic emotional expression from the video. It itself captures the images wherever the faces can be visible in the video, and then code them into the basic emotions. They have stated that this provides savings in processing time which is important for real-time applications. In their work, they have used Cohn and Kanade's DFAT -504 dataset to train and test their system. It consists of 100 university students ranging in age from 18 to 30 yrs. The proposed system scans all possible 24*24 pixel patches in the image and classifies each as face vs non-face. It consists of a cascade of classifiers, in which each contains a subset of filters reminiscent of HAAR Basis function, that can be computed very fast at any location and scale in constant time. The located faces were rescaled to 48*48 pixels so that it can match with the dataset. Then they have converted the images into a gobar magnitude representation, using a bank of gobar filters at 8 orientations and 5 spatial frequencies. And they have classified the facial expression based on support vector machines (SVM's). It is done in two stages, first, the SVM performed binary decision tasks. Seven SVM was trained for each emotion classification. The emotion category decision with the maximum margin for the test. Linear, polynomial, and RBF kernels with Laplacian and Gaussian basis functions were explored. By

this, they have made a user-independent fully automatic real time coding of basic expressions is an achievable goal with present computer power.

Hong-Wei Ng et al [2], they present the techniques used for submissions to the 2015 Emotion Recognition in the Wild (EmotiW) contest, for the sub-challenge of Static Facial Emotion Recognition in the Wild (SFEW). A contest namely EmotiW and challenge SFEW, follow same way of approach for the seven basic expressions. pictures are taken from movies, in a semi-automated manner, through a system supported subtitles. The difficult features of SFEW are twofold. First feature of SFEW is imaging conditions that are close to the human life, as well as low and uneven illumination, low resolution, occlusions, non-frontal head-poses, and motion blur. Second, the dimension of the dataset is comparatively small, that makes it troublesome to train large-scale models and so, is prone to overfitting. Approaches supported deep learning and people using CNNs are successful at image-related tasks due to their ability to extract sensible representations from data. Judging a person's feeling will be troublesome even for humans sometimes, due to refined variations in expressions between the additional nuanced emotions such as sad and fear. As a result, efficient features, finely-tuned and optimized for this specific task are measure of great importance so as for a classifier to create sensible predictions.

It comes as no surprise that CNNs have worked well for emotional classification, as proved by their use of a variety of state of art algorithms for this task, additionally winning competitions, particularly previous years EmotiW challenge. However, because of the tiny dataset size for the EmotiW 2015 image based SFEW challenge, it is easy for advanced models like CNNs to overfit the data. To solve this problem of coaching a high capacity classifier on tiny datasets, previous works in this area have resorted to transfer learning across tasks, where the weights of the CNN measure initialized with those from a network trained for tasks before fine-tuning them with the target dataset. This approach has consistently achieved higher results, compared to directly train the network on the tiny dataset and that is what they tend to adopt during their paper additionally.

Peter Burket et al [3], have proposed architecture that is independent of any hand-crafted feature extraction and performs better than the earlier proposed convolutional neural network approaches. The presented approach uses Artificial Neural Networks (ANN). ANN differs from others by they are trained on the data with less need for manual inference. Convolutional Neural Networks are a special kind of ANN and have been shown to work well as feature extractor when using images as input and are real-time capable. Google Net is a deep neural network architecture that relies on CNN's. The proposed Convolutional Neural Network architecture consists of four parts. The first part automatically pre-processes the data. This starts with Convolution 1, which applies 64 different filters. and then comes the Pooling 1 layer to down sample the images and then normalized by LRN and then Parallel Feature Extraction Block. They are the core of the proposed architecture. data extracted by these blocks are forwarded to a completely connected layer, that uses them to classify the input into the various emotions. The key structure in their architecture is the Parallel Feature Extraction Block (FeatEx). This application uses different sizes of filters to detect faces. The paths are combined for a more diverse representation of the input. Using this block twice yields good results. Here they have used the MMI dataset which contains over 2900 videos and images of 75 persons and the CKP dataset where 210 persons (in age of 18 to 50) have been recorded depicting

emotions. Hence a neural network that has little computational effort compared to the current state of the art CNN architecture has been proposed.

Sreelakshmi et al [4], they proposed real time facial features identification method using active shape model (ASM) and look based mostly feature extraction. This methodology is used to enhance the driver safety by recognizing the expression of the driver. First, we tend to discover human faces using the quality Viola-Jones algorithmic rule. Based on the face region that has been detected the positions of lip region, region between eyebrows and region near to the lip corner known using ASM. Once extracting the positions of these regions, a kirsch compass filter applied to the face image for edge detection. In next step we tend to extract the edge data from the regions obtained from the ASM. the edge data in these regions is employed as features to classify the facial expression using multiclass SVM classifier. This method can categorize the expressions into four standard classes namely neutral, smiling, surprise and disgust.

The idea is the real time identification of facial emotion recognition based on appearance features. They have developed a way that may identify facial expression by fitting a model to the face. It extracts options from the kirsch mask convoluted candidate image. The system is tested on each real time and offline pictures. It will classify neutral, smile, surprise and disgust expression with a mean accuracy of 84.5%. The accuracy of the algorithmic rule is greatly influenced by the accuracy of ASM fitting. ASM fitting fails in conditions once the face is partially occluded, when it's extremely tilted, for individuals with sun glass or spectacles with thick frame and for individuals with frontal hair falling on forehead. The feature extraction does not provide higher result once the person is thick bearded. The misclassification rate for anger expression is a lot of for this method. A potential reason for this may be that individuals express their anger in numerous ways which can be troublesome to characterize only the required region.

Facial expressions additionally to verbal communication, convey our mental and emotional state. Facial expressions are result of movement of many muscles on face that make lines, wrinkles, furrows, deformation of some facial expression. It's a comparatively straight forward task for human being to look at human faces and analyse the expressions on the face. It provides data of the emotional state of a person that may be a vital input to own. However, making machines that may discover and analyse human expressions is not a trivial task.

Gerard Pons et al [5], stated that Although recent methods achieve close to human accuracy in controlled scenarios, the recognition of emotions in the wild remains a challenging problem. They state that used of ensembles of CNN is more efficient than the individual CNN classifiers. The following two statements proves the above statement: (i) in CNNs design there is an adjustment of parameters which allows diversity and complementary in classification results, (ii)then final classification rule assembles result of committee. In this paper they propose to improve the assembling of the committee by introducing supervised learning on the ensemble computation. They trained a set of 72 CNNs using a baseline architecture of 4 layers, with diverse configurations including several parameters as well as different initializations from pre-trained models using different databases. With the posterior-class probabilities of these individual networks, formed a hierarchical committee based on a deep CNN. Second, they repeated this process with a set of 64 CNNs using the VGG-16 as a baseline architecture. Their proposed approach was assessed on the SFEW2.0 database for the EmotiW 2015 subchallenge,

and its performance was compared to the most used committee strategies, majority voting and average rule, the exponentially weighted average VAexpoWA, and two non-linear methods, SVM and NN, obtaining the best results in terms of accuracy for both experiments.

Shishir Kumar et al [6], proposed that Facial Action Coding System (FACS) could be a helpful structure that classify the human facial actions by their advent on the face using Action Units (AU). Facial and emotional expressions are the foremost vital nonverbal ways which express internal emotions and intentions. AU is one among forty-six minor components of visible facial motion or its connected type changes. Facial expressions have worldwide meaning and these emotions are accepted from so many years and it absolutely was the most reason to pick facial expressions for the research. Facial Expression Recognition (FER) could be a difficult task in machine learning with a wide-range of applications in aid, HCI, and gaming. Emotion recognition is difficult because of many input modalities, have a major role in understanding it. The mission of recognizing of the feelings is generally tough because of 2 main reasons: 1) There's not mostly accessible information of trained and input pictures and 2) Classifying emotion couldn't be simple based on whether or not the input image is static or evolution frame into a face expression. They proposed a deep learning technique within the context of emotional recognition so as to classify emotion labels from the pictures. Several ways and analysis have been developed during this regards, however most current works are appeared that specialize in hand-engineered features. currently a day’s because of quantity and variety of datasets, deep learning is changing into mainstream thought techniques in all computer visions tasks. Conventional convolutional neural systems have an interesting constraint that they simply handle spatial image. The fundamental commitment of this work is to show the Spatio-worldly development of outward appearances of a person inside the pictures utilizing a Recurrent Neural Network (RNN) that embedded with a Convolutional Neural Network (CNN) in form of CNN–RNN design.

Amr Mostafa et al [7], have used the concept of emotion recognition by facial features using recurrent Neural Networks. Affective computing is the branch of human-computer interactions which depending on those activities tries to use various machine learning algorithms to predict human emotional states to improve the computer usability interface. Emotion recognition generally has been studied extensively in the recent literature using different datasets but none of them used the discrete dimension representation of emotions as represented in the BioVid Emo DB. The BioVid Emo DB is designed to contain high-quality data of induced five discrete emotions. It is designed by eliciting an emotion using certain film clips, as it has been proven to be the most effective rather than pictures and music. In the experiment, a total of 94 participants were examined and none of them had affective or related disorders. They form the following age groups: 18-35 (N = 35; 16 men / 19 women); 36-50 (N = 31; 13 men / 18 women); 51-65 (N = 28; 15 men / 13 women).

The emotion recognition field was the main point of researchers in the psychology. Emotion recognition system passes by three main stages. In the first stage, the face is being located and detected in each frame of the video sequence from the dataset. Then different features are being extracted from the facial regions of the video sequences as an input to the machine learners. Two categories can be used for features in the feature extraction phase: geometric-based and appearance-based features. Geometric-based features are extracted from a set of computed landmarks consisted of 68 2D facial points describing eyes, nose, and mouth regions moreover

the shape of the detected facial region at a frame level. The Euclidean metric is used to calculate the ordinary distance between each facial point and the centre of gravity (COG) of the total points with an additional 18 distance. Then the total sequence of the distances is statistically grouped to form the feature vector to be fed to the machine learning model. The experimental findings suggest that emotion discrimination from video features is feasible. Especially discriminating between anger and disgust expressions. Overall, utilizing this research work after proving its validity will be an effective tool in Health Agencies or in physical infirmaries, mental rehabilitation centres such as nurse to patient approaches to achieve short term and long-term treatment goals of the patients coping mechanism to depression, fear, disgust brought about by the physical and mental illness they are in.

Ali Sharif Razavian et al [8], have carried out paintings on CNN capabilities.

Deep learning. “How well do you have watched it would paintings in your pc vision trouble?” maximum in all likelihood this question has been posted on your institution’s coffee room. And in response, someone has quoted current success testimonies and a person else professed scepticism to undertake the CNN illustration to address the problem of photo category of items and scenes. The gadget should assign (potentially multiple) semantic labels to a photo. Don't forget in contrast to object detection, the item image category calls for no localization of the objects. The CNN representation has been optimized for the item picture type mission of ILSVRC. The result of different retrieval strategies applied to five datasets. Spatial search is the handiest used for the primary three datasets that have samples in one-of-a-kind scales and locations. For the other datasets, we used the identical jittering as explained in a sec.

Yue sun et al [9], have accomplished work on outward appearance Recognition Based on Arousal-valence Emotion Model and Deep Learning Method. With the improvement of computerized reasoning, outward appearance acknowledgment has been widely concentrated in human-PC communication. Outward appearance is the human's primary method to communicate feelings. Analyst Mehrabian characterized the notable 7%-38%55% standard for feeling demeanor, that is, the feeling is communicated by 7% language, 38% voice, and 55% outward appearance. As of late, outward appearance acknowledgment is applied in numerous zones, for example, human-PC association, driver weariness screen, and patient consideration in the emergency clinic.

Investigations are led on the AVE20130, RECOLA, and NVIE datasets in this paper. Face location utilized HAAR include classifier with OpenCV. The picture pre-preparing strategy is given. CNN model expects pictures to be changed to a fixed size (224×224). Every crude picture is sent into CNN and scholarly the upgraded highlights. In the investigation, the model of CNN we utilized is from

MATCONVNET, MATCONVNET is a MATLAB tool stash. So as to diminish. Outward appearance acknowledgment dependent on excitement valence feeling model is at present uncommon in residential research. Profound learning is received in this paper. It has a critical preferred position in outward appearance acknowledgment. The future research will concentrate on the SoftMax layer, the last layer of CNN will be supplanted with a projection layer.

3. REAL LIFE APPLICATION

Human emotions are very complex to understand. Facial emotion recognition reveals complex mental states and indicate the internal states that are crucial in social interaction. This technology can be used in Psychotherapy to understand the patients and to get a better picture of their condition. Emotion recognition in psychiatric patients is essential in the matter of both their treatment and diagnosis.

In order to predict the mental state of the patient, psychiatrists can use this technology to keep track of a patient mental state and monitor them by continuously getting updated by their emotions each second. In particular, fear recognition will be higher in residents with high anxiety and hostility scores, and enthusiastic qualities will be connected with the recognition of disgust. It is considered that accurate emotional recognition and analysis may have a significant impact on the patient-doctor relationship. It is valuable to know how the ability of facial emotion recognition would impact the quality of patient-clinician interaction.

4. INDIVIDUAL CONTRIBUTION

18BCE0541- Deciding the best model by comparing the details for the project and implementation which involves taking the existing neural network for recognising emotions and capture the multiple faces at the same time from the live video capture and detecting the emotions.

18BCE0624- Keeping track of emotion at every second. Record every change in the emotion and display the emotion for every second. Alert the user if the emotion of the person reaches the threshold via email.

18BCE0595- Make the login form and added instructions page which has guidelines to be followed while using the application.

18BCE0606- Making the use of Tkinter and necessary python modules to achieve a user interface which is not only fast in response but also very easy to understand and use (collect the details of the user).

5. TOOLS AND TECHNOLOGIES USED

- Computer vision: Computer Vision algorithms perform image processing to extract features and use them for classification for example, we can think about them that they are handy crafted.
- Deep learning: We use deep learning in the form of Convolutional Neural Networks to perform facial recognition. Deep learning model learns them self and extracts features automatically.
- DCNN (Deep Convolutional Neural Network): the DCNN model, the convolution and sampling layers are combined into a single layer. Based on the already trained network, greatly improve the image recognition rate.

- GPGPU (General purpose graphic processing unit): Reduce the complete load on the CPU for the real time use of the video capturing for the face recognition part.
- SMTP (Simple Mail Transfer Protocol):it helps in controlling the mailing server to mail from an email account with the help of a software
- Tkinter Python Module: The following module of python is to be used for the interface part of the project

6. PROPOSED SYSTEM PROCESS FLOW

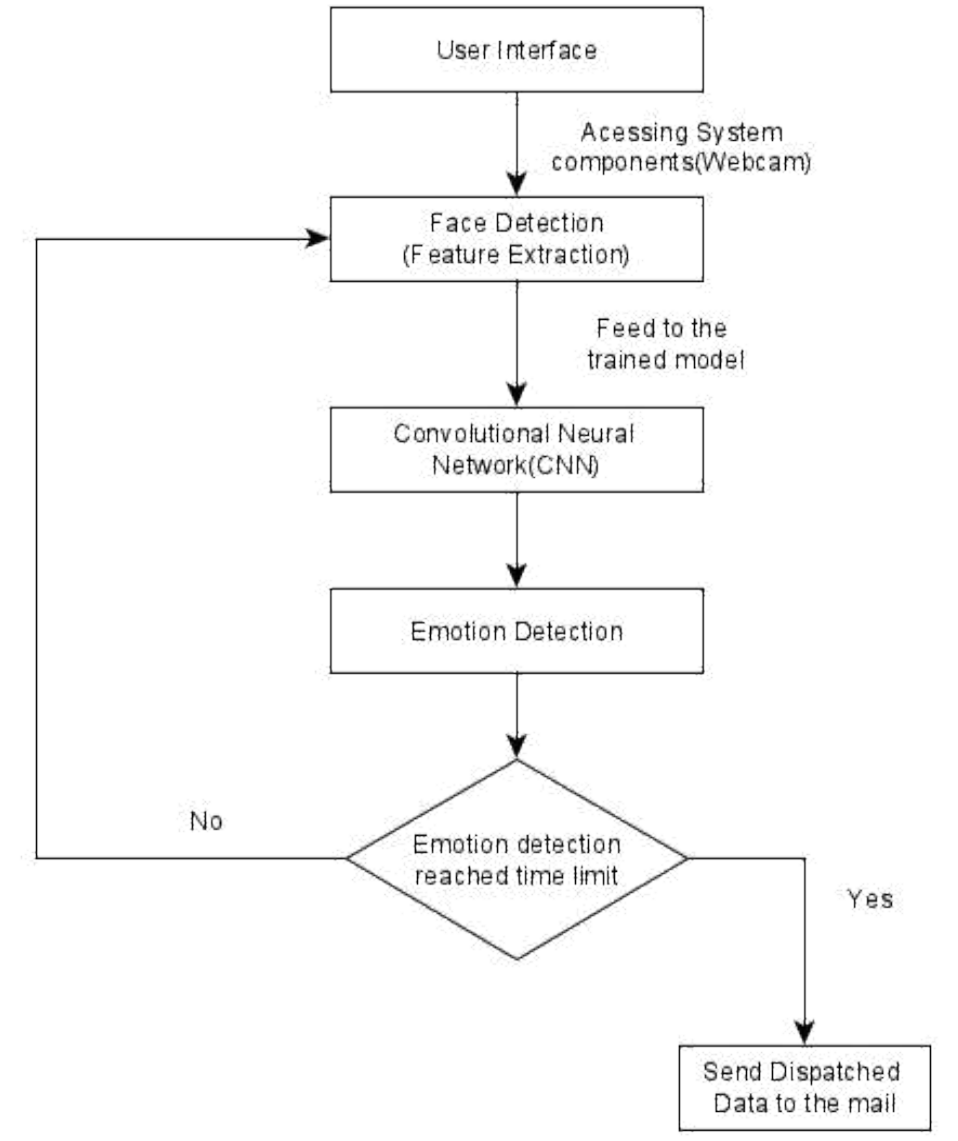


Fig 6.1- Proposed System

7. WORKING METHODOLOGY

The Aim of the project is to predict the emotion of the person by their facial expression and swap the appropriate emotion in place of the face. There are 7 types of human emotions that are recognised universally namely anger, happiness, disgust, fear, sad, surprise and neutral.

Preparing the dataset: It involves the preparation of dataset upon which the training algorithm work. The Model we will be using contains different datasets of faces for example 'FER2013' dataset and some images followed by applying it to convolution neural networks. This model will help to recognize emotions of the person.

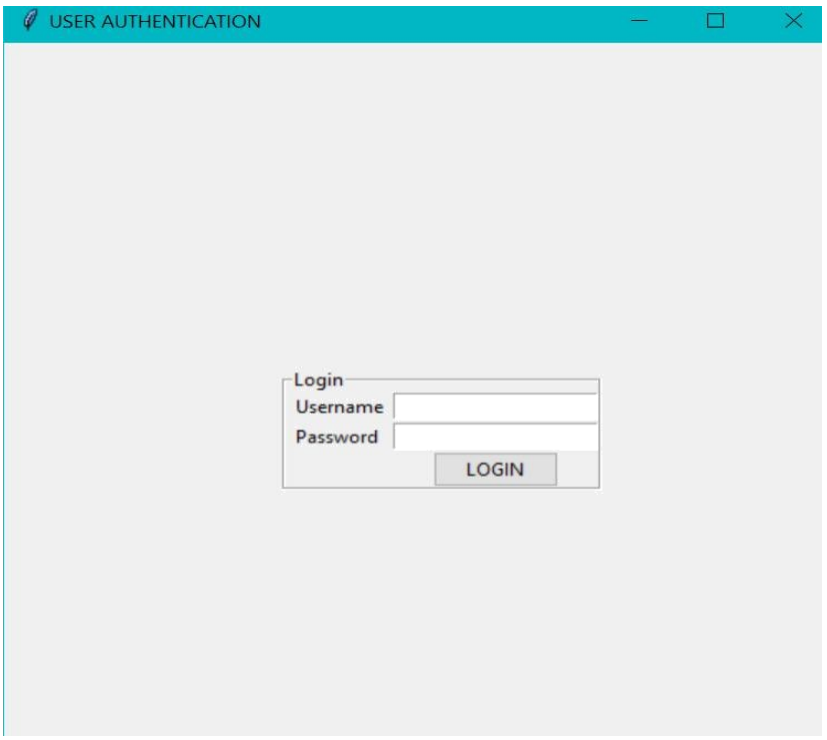
Detecting faces: This is done by using HAAR course work in OpenCV by recognizing the appearances, the image is changed to grey -scale and is resized to a similar size as the images in dataset. A HAAR course is fundamentally a classifier which is used to detect specific objects from the source.

Swapping emotions in place of face: This is the important step to recognize emotion and it involves placing the appropriate emotion over the face of the person according to their emotion. The HAAR cascade function returns the coordinates of the face detected and these coordinates can be used to place the emotion of the person at the correct place.

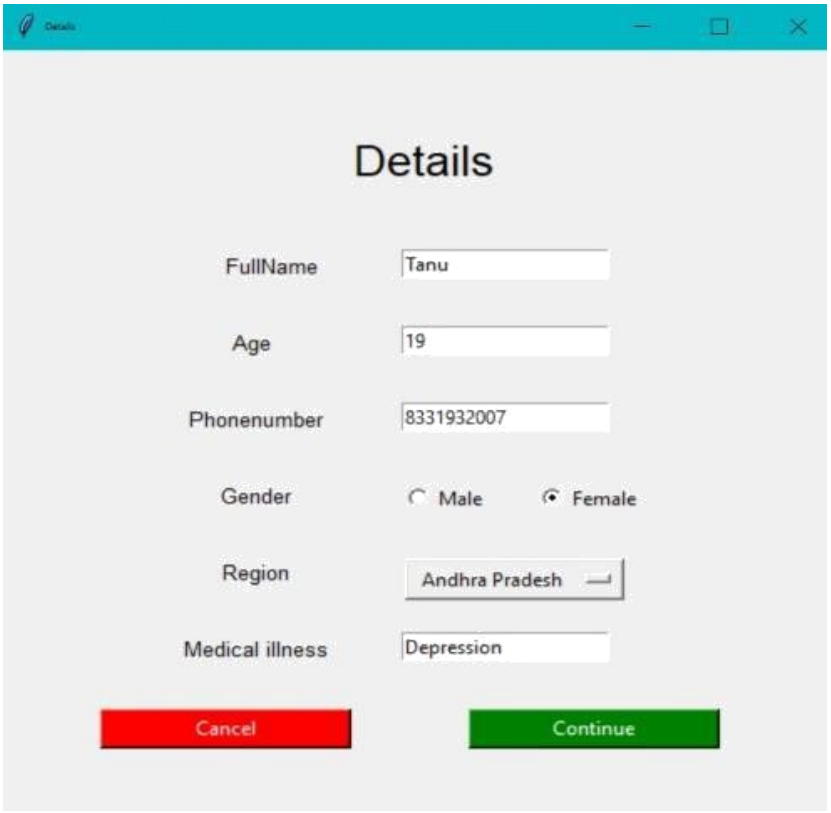
Parsing the data and emotions: We will parse the data frame by frame and the output will be in the form of text file containing emotions frame by frame. Now we will set a limit for certain emotion like fear for which it sends the data if the limit is crossed.

8. IMPLEMENTATION RESULTS AND USER INTERFACE

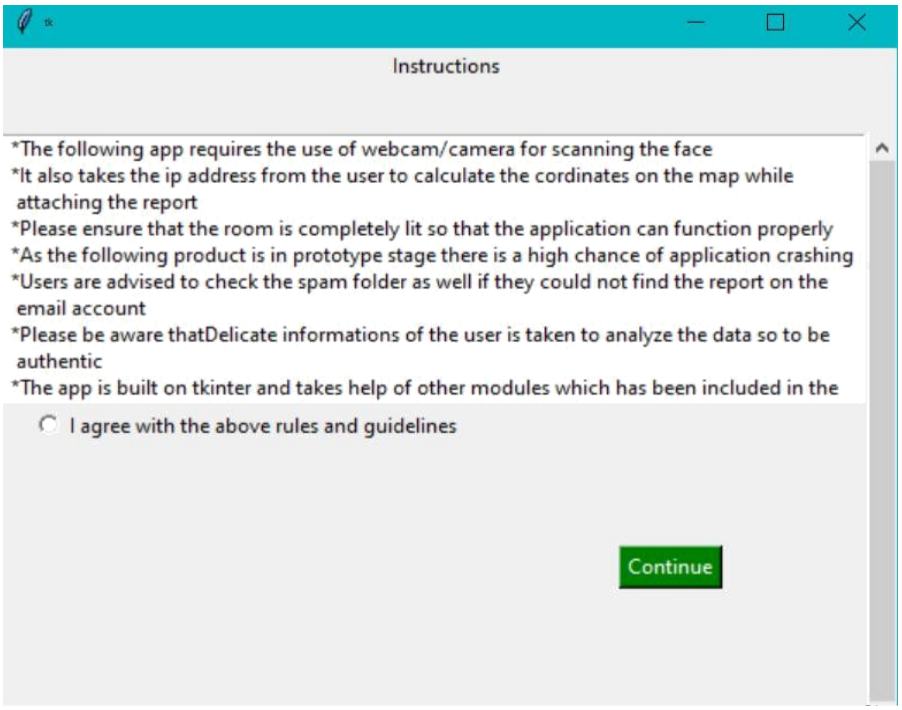
This is login form where only admin has the access to use this software. The admin has to type username and password to login to the application.



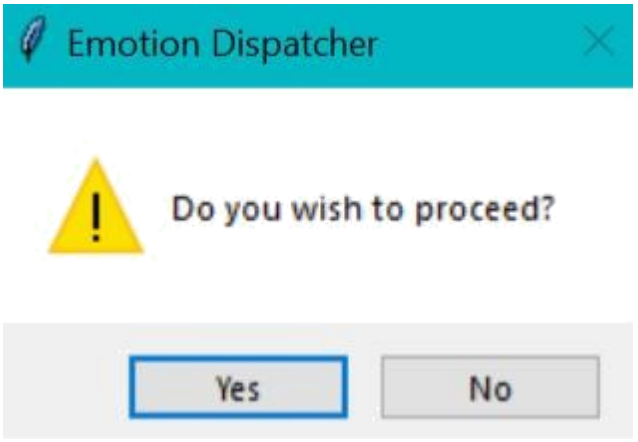
This is the registration form to fill the details of the user and the details will be sent to the mail after detecting their emotions. This can help the companies to visualize them and help to improve their market strategy. The registration form contains the following components: Name of the person, Age, Phone Number, gender, region and their medical illness details.



This page consists of instructions to be followed while using this application.

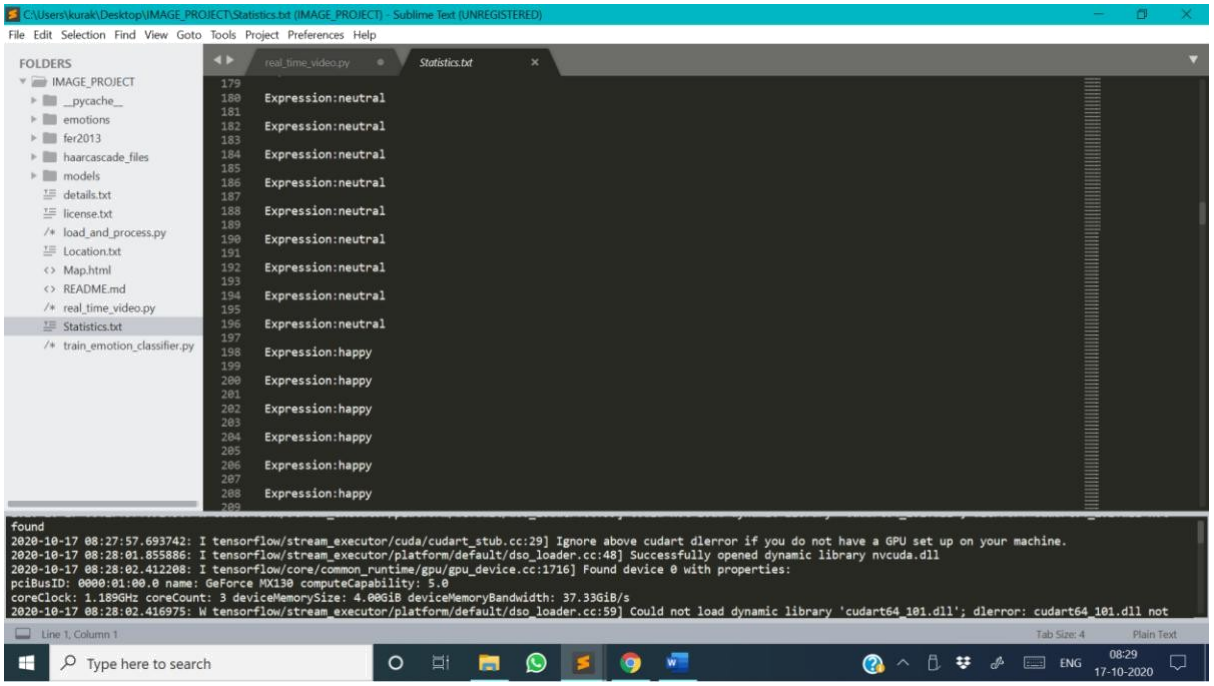


To detect the face, asking the user's permission to allow the access to the webcam.



The program will create a window to display the scene capture by web camera and a window representing the probabilities of detected emotions.

The expressions at each frame are recorded and stored in a text file.

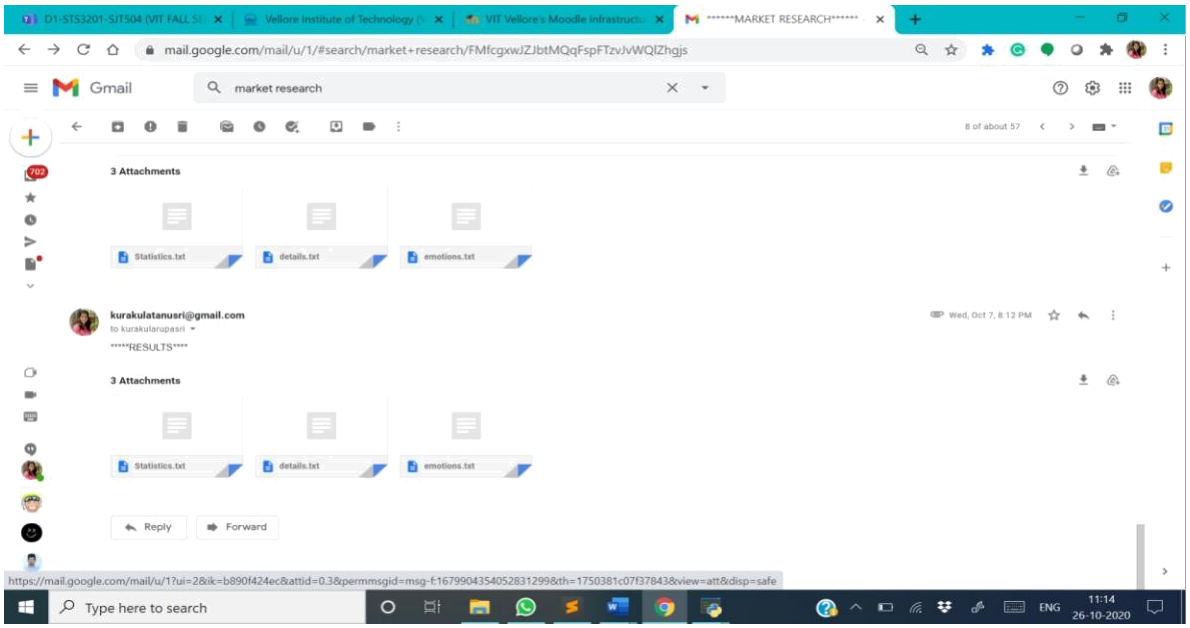


After a certain period of time detected emotions and the details of the user is dispatched through mail. Three files will be dispatched at every detection.

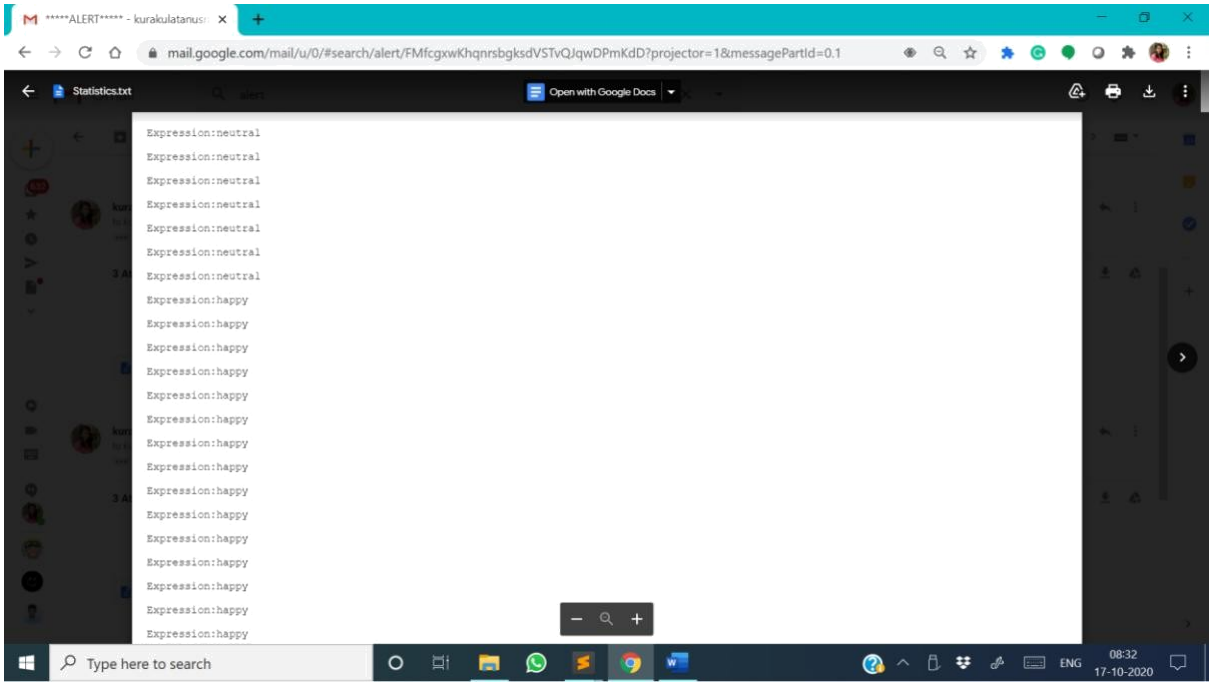
Statistics.txt → This file contains the emotions of the user at each frame

Details.txt → This file contains details of the user which is filled in the form.

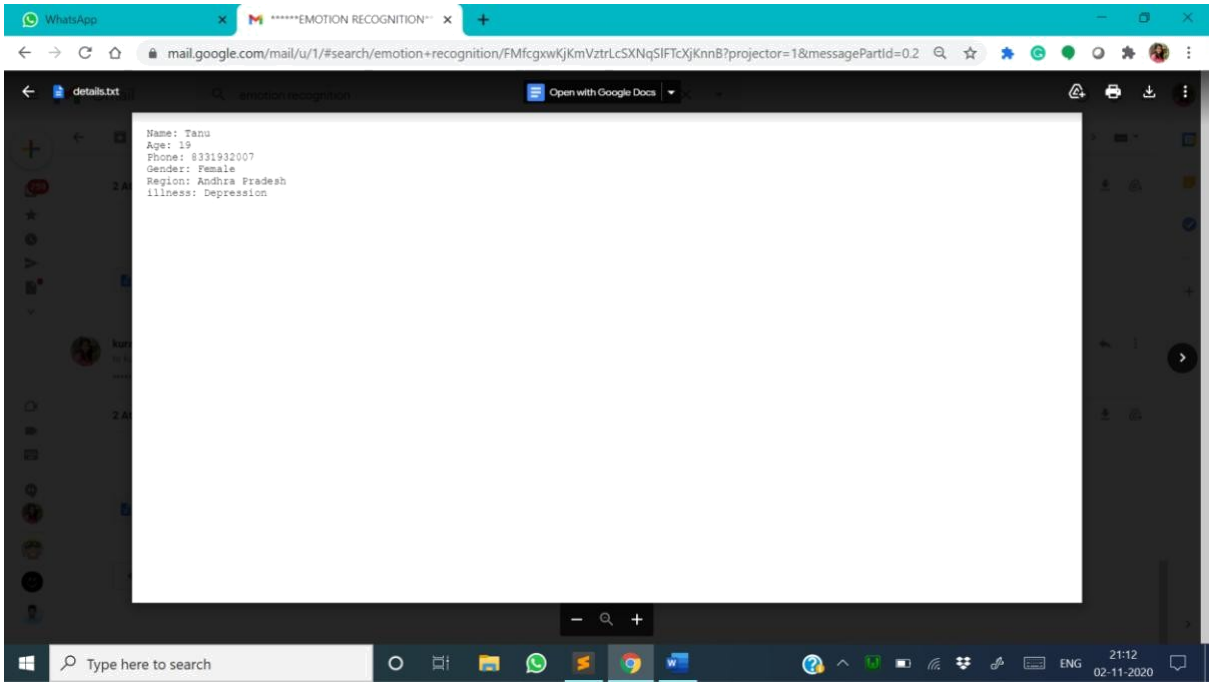
Emotions.txt → This file contains counter of each emotion.



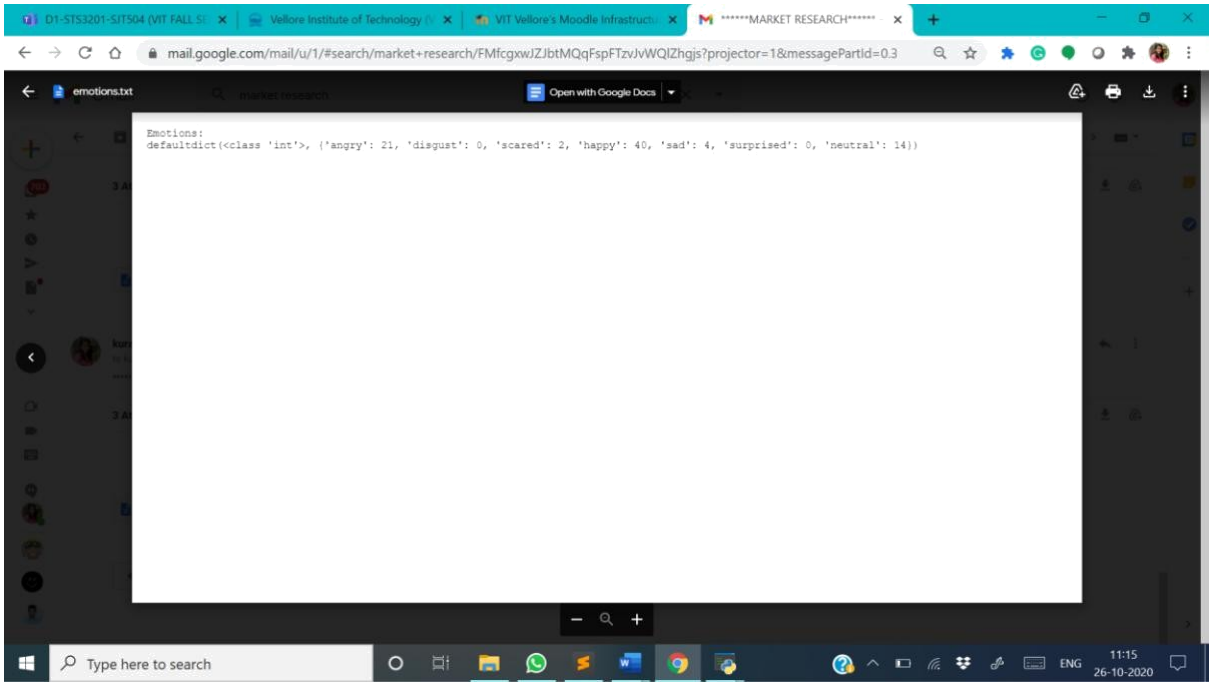
Statistics.txt: Emotions of the user at each frame.



details.txt: Details of the user



Emotions.txt: Counter of each emotion.



9. INTERFACE VALIDATION

Neilson's heuristic principles:

- **Visibility of the system status**: One knows when the face recognition takes place as the screen is popped up in the window to recognize the emotions through web cam. The probabilities of expressions is shown at each frame.
- **Match between the system and the real world:** The application works like any other form fill up app to get the information and later give feedback based on your emotion. As the application can easily be used in variety of systems it gives the user a feeling of similarity.
- **User control and freedom:** The user is allowed to cancel the task if he is not willing to agree with the guidelines.
- **Consistency and standard:** Consistency and Standards involves the consistent display of information. This requires the system dependably use standard formats, fonts and spacing. The consistency is maintained in the interface
- **Error prevention:** At every page in the interface cancel button is provided to undo the action performed by the user.
- **Recognition rather than recall:** Information once taken will not be asked again instead will be used in all the required spaces
- **Flexibility and efficiency of use:** The application can easily work on laptops and desktop with minimalistic graphical setting and a webcam. The entire application is pre-modelled therefore won't take much time to load or move it to different systems.

- **Aesthetic and minimalist design**: Used the most widely used colour pattern for the application with minimum animation to render it easily. Made every application window of same size and made the fonts and button bolder.
- **Help users recognize, diagnose and recover from errors**: Specifies the error on the spot with an alert message and specify the error in detail. Try to make the errors sound less technical so that everyone can understand it.
- **Help and documentation:** The application will have a guidelines page containing the documentation.

10. COMPARITIVE ANALYSIS

Comparison of models:

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
VGG16	528MB	0.713	0.901	138,357,544	23
InceptionV3	92MB	0.779	0.937	23,851,784	159
ResNet50	98MB	0.749	0.921	25,636,712	-
Xception	88MB	0.790	0.945	22,910,480	126
InceptionResNetV2	215MB	0.803	0.953	55,873,736	572
ResNetXt50	96MB	0.777	0.938	25,097,128	-

VGG-16

VGG-16 is a less difficult design model since it's not utilizing many hyper boundaries. It generally utilizes 3 x 3 channels with a step of 1 in the convolution layer and uses the SAME cushioning in pooling layers 2 x 2 with a step of 2.

GoogLeNet

The champ of ILSVRC 2014 and GoogLeNet engineering is otherwise called the Inception Module. It goes further in equal ways with various open field sizes and it accomplished a best 5 blunder rate with 6.67%.

ResNet (2015)

The victor of ILSRVC 2015, it likewise called as Residual Neural Network (ResNet) by Kaiming. This engineering presented an idea called "skip associations". Ordinarily, the info lattice figures in two straight change with ReLU actuation work. In Residual organization, it straightforwardly duplicates the info network to the subsequent change yield and entirety the yield in the last ReLU work.

Inception:

The Inception module registers numerous various changes over a similar information map in equal, interfacing the outcomes into a solitary yield. For each layer, it does a 5x5 convolution, 3x3 convolution, and max pooling, each conveys distinctive data, which obviously is computationally exorbitant. Consequently, the creators of Inception chose to conquer this issue by presenting the measurement decreases.

MobileNet:

The main idea behind MobileNet is lighter deep neural networks are assembled using depthwise separable convolutions. Usually, in the convolutional layer, the convolution kernel or filter is applied to all of the channels of the input image, by calculating the weighted sum of the input pixels with the filter and then slides over the next input pixels across the images. This regular convolution is used by MobileNet only in the first layer. The next layers are the depthwise separable convolutions which are the combination of the depthwise and pointwise convolution.

Xception:

Francois Chollet proposed Xception Model. Xception is an extension of the inception Architecture that replaces the standard Inception modules with depthwise Separable Convolutions. Xception is a convolutional neural network that consists of 71 layers. This allows us to load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pre-trained network can classify images into 1000 object categories, such as vehicles, keyboards, mouse, and many animals. Thus, the network has learned rich feature representations for a wide range of images. The network has an image input size of 299-by-299.

11. CONCLUSION AND FUTURE SCOPE

Facial emotion recognition is an emerging field now-a-days. Non-verbal communications like facial expressions are used in many applications in human computer and interaction which is used to convey the facial emotions. There is lot of complexity and variability involved in recognizing facial emotions. This project proposed a new method for facial recognition and its various applications in real time scenario. Predicting emotions give us accurate information of how the user feels. So this proposed method helps the psychiatrists to analyse their patients effectively. We believe that this proposed method has given promising results in detecting the emotions and dispatching the data.

12. REFERENCES

[1]. M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.," *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Madison, Wisconsin, USA, 2003, pp. 53-53.

[2]. 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.

[3]. Burkert, Peter & Trier, Felix & Afzal, Muhammad Zeshan & Dengel, Andreas & Liwicki, Marcus. (2015). DeXpression: Deep Convolutional Neural Network for Expression Recognition.

[4]. Sreelakshmi, C. and Kutty, K., "Vision Based Face Expression Recognition," SAE Technical Paper 2015-01-0218, 2015

[5]. 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, 1 July-Sept. 2018. doi: 10.1109/TAFFC.2017.2753235

[6]. Neha Jaina, Shishir Kumar , Amit Kumar , Pourya Shamsolmoali, Masoumeh Zareapoor, “Hybrid deep neural networks for face emotion recognition”, 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. doi: 10.1109/ICCES.2018.8639182

[8]. A. S. Razavian, H. Azizpour, J. Sullivan and S. Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, 2014, pp. 512-519. doi: 10.1109/CVPRW.2014.131

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

doi: 10.1109/ICCTEC.2017.00022

[10] Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

13. APPENDIX

Link to ppt:
https://drive.google.com/file/d/1dtzDT-21YBs7SqF5k5KP4rX1D8z2-jY_/view?usp=sharing

Link to pre-recorded demonstration video:
https://drive.google.com/file/d/1Y8BppE7fK_1ZInUa8aOf09QuEvg8CJUM/view?usp=sharing

Link to access source files:
https://drive.google.com/file/d/1Eo29fLSgTxhNiFI-RkKIDgr9NbOSCrBg/view?usp=sharing