**WALLPAPER SUGGESTION THROUGH**

**EMOTION RECOGNITION**

RAHUL KOCHAR (19BCE0894) ANVI MITTAL (19BCE0906) ARYAN PATRO (19BCE0904)

*Vellore Institute of Technology, Vellore Institute of Technology, Vellore Institute of Technology,*

*Vellore Vellore Vellore*

***Abstract*—Human Emotion Recognition is proposed to be a definite system wherein the person’s emotions can be detected using facial expressions, speech, gestures, gait, and hand movements. However important has emotion recognition become in this era, it isn’t open in this world now. Over the last years, face recognition and automatic analysis of facial expressions has become one among the foremost challenging research areas within the field of Computer vision and has received a special importance. Emotion plays a crucial role to understand the feeling of each person’s figure clearly about a person's personality. Facial and emotional Expression recognition is an important technique as by using Facial Emotion Recognition, businesses can process images, and videos in real-time for monitoring video feeds or automating video analytics, thus saving costs and making life better for their users. Hence, we would like to bring it together with emotion detection using OpenCV face recognition classes via webcam and an open-source library for data flow. In our project, specifically, we have used Emotion Detection to suggest the user wallpapers based on his/ her current emotions.**

***Index Terms*—Emotion Recognition, computer vision, OpenCV, CNN, Classification.**

1. INTRODUCTION

acial expressions are important in facilitating human

communication and interactions. Also, they are used as an important tool in behavioral studies and in medical rehabilitation. Facial image based mood detection techniques may provide a fast and practical approach for non-invasive mood detection. The purpose of the present study was to develop an intelligent system for facial image based expression classification using committee neural networks.

The objective of the projects is to train a model of a convolution neural network. Broadly it is efficient emotion recognition by using the Keras model and convolutional neural network for the project, leaving the scope of tuning of hyperparameters in the future. Implementation of Red, green, blue scaling instead of grey scaling and high epoch value for better accuracy and detection irrespective of the background. Seven types of emotions recognized followed by the algorithm, broadening the range of the emotion that can be recognized followed by wallpaper suggestions.

The Project aims to suggest wallpapers depended on real time emotions by taking one’s picture to detect and standardize the emotion and suggest wallpapers as per their emotions. Hereby, taking the emotion detection and thereafter taking up their mood wallpapers, hand in hand.

Facial expression images were obtained from the user input after successful sign in followed by authentication of the user. Then detected emotions will be shown on the screen. Later the user will see different emotions like, neutral, happy, surprise, anger, fear, disgust, and sad, and thereafter after clicking on any particular emotion or better the detected one, user will be able to see wallpapers related to that emotion specifically, and will be able to download the wallpapers too.

II. LITERATURE SURVEY

i**) Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information**

As computers become able to recognize and react to human nonverbal communications such as emotions, the interaction between humans and computers becomes more natural. Several approaches have been proposed for recognizing human emotions based on facial expressions and language, but these two modality and other modality are used to improve the accuracy and robustness of emotion recognition systems. The work to combine is relatively limited. In this article, we will analyze the strengths and limitations of a system that is based solely on facial expressions and acoustic information. It also describes the two approaches used to merge these two modality: decision-making and functional integration. Based on the database recorded by the actress, four emotions were categorized: sadness, anger, happiness, and neutrality. By using markers on her face, she combined motion capture with simultaneous voice recording to capture detailed facial movements. The results show that facial expression-based systems performed better than systems based solely on the acoustic information of the emotions being seen. The results also show the complementarity of the two modalities, and the integration of these two modalities can significantly improve the performance and robustness of the emotion recognition system.

**ii) Face Recognition Based on Convolutional Neural Network**

Face recognition is very important for real-world applications such as video surveillance, human-machine interaction, and security systems. Compared to traditional machine learning approaches, deep learning-based methods show superior performance in terms of image recognition accuracy and processing speed. This white paper proposes a modified convolutional neural network (CNN) architecture by adding two normalization operations to the two layers. Normalization operations, which are batch normalizations, speed up the network. We used the CNN architecture to extract characteristic facial features and used the Softmax classifier to classify faces in fully connected layers of the CNN. In the experiment part, Georgia Tech Database showed that the proposed approach has improved the face recognition performance with better recognition results.

**iii) Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network**

Emotion recognition using facial expression is very much necessary these days. Different kinds of emotions reflect different definitions. Facial emotion recognition plays a major role in driver warning systems, it can also play an important role in shopping malls to predict unusual activity like terrorist attacks, robbery and much more. Predicting the suicidal tendency of a person also can be done using facial emotion recognition. An automatic facial emotion classification system is proposed in this paper using the Convolution Neural Network (CNN) with the features extracted from the Speeded Up Robust Features (SURF). 91. Accuracy is achieved with the proposed model. This helps track human emotions with facial expressions.

III. EMOTION RECOGNITION SYSTEMS

**2.1 Emotion recognition by speech**

Several approaches have been reported for recognizing emotions from voice. Most researchers use global suprafix / prosodic features as auditory cues for emotion recognition to calculate speech-level statistics. For example, the pitch contour and energy mean, standard deviation, maximum and minimum values ​​of an utterance are widely used features in this regard. Dellaert et al. We try to classify four human emotions using space-related features. They implemented three different classifiers: Maximum Likelihood Bayes classifier (MLB), Kernel regression (KR) and nearest neighbor method (KNN). Roy and Pentland used a linear Fisher classifier to classify emotions. In a short sentence, they recognized two types of emotions: approved or disapproved. They performed some experiments on the features extracted from the pitch and energy measurements and obtained an accuracy of 65% to 88%.

The main limitation of these acoustic features at the global level is the inability to describe dynamic changes along with utterances. To remedy this, for example, dynamic changes in speech emotions in spectral changes can be tracked at the local segmental level using short-term spectral features. Using the 13-mel frequency cepstrum coefficient (MFCC), we trained a hidden Markov model (HMM) and recognized four emotions. Nwe et al. Using 12 mel-based audio signal power coefficients, we train a discrete hidden Markov model and classify 6 typical emotions.

The average accuracy of both approaches was 70-75%. Finally, other approaches use language and discourse information to examine the fact that some words are highly correlated with a particular emotion. In this study, prosodic information is used not only for the duration of voiced and unvoiced segments, but also for acoustic properties.

**2.2 Emotion recognition by bimodal data**

Relatively few efforts have focused on implementing emotion recognition systems that use both facial expressions and acoustic information. De Silva et al. We have proposed a rule-based audiovisual emotion recognition system that integrates the output of a monomodal classifier at the decision-making level. The audio used the prosodic function, and the video used the maximum distance and speed between six specific viewpoints. A similar approach was used by Chen et al. The dominant modality from subjective experiments performed in was used to resolve discrepancies between the outputs of monomodal systems. It is concluded that using both modality together would improve system performance.

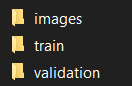
We have proposed a multimodal system that considers not only audio and visual information but also the heat distribution detected by an infrared camera. They claim that infrared images are not sensitive to light. This is one of the main problems when capturing facial expressions with traditional cameras. They used a database recorded by a female speaker who read one word in five emotional states. At the decision-making level, they integrated these three modality with empirically determined weights. Using the three modality together improved system performance.

A bimodal emotion recognition system has been proposed for the recognition of six emotions in which audiovisual data is fused at the feature level. They used prosodic functions from audio and the position and movement of facial organs from 206 videos. The best characteristics of both monomodal systems were used as inputs to the bimodal classifier. They showed a significant improvement in performance from 69.4% (video system) or 75% (audio system) to 97.2% (bimodal system). However, because we are using a small database with only 6 clips per emotion, the generalizability and robustness of the results should be tested on larger datasets.

All of these studies show that the use of multimodal information can improve the performance of emotion recognition systems. However, it is not clear which method is best for merging these modality. This white paper addresses this open issue by comparing system performance decision-making with functional-level integration techniques.

IV. TRAINING PHASE

So in any deep learning model we start by training our model. Our goal here is to distinguish between pictures for various emotions where the images show the different kinds of emotions of a person. So for training we need two datasets in each folder of training, validation and testing.



So we start off by organizing our data in google drive to which we can link our notebook and extract data as we wish. We keep our Jupyter notebooks in a folder called Colab notebooks and we keep all the data in a folder called data.

V.METHODOLOGY

The convolutional layer is a Convolutional network that performs most computational tasks Lift heavy objects. The main purpose of the folding layer is Features are extracted from the input data that is an image. Convolution preserves the spatial relationship between pixels Learn the features of the image using the small squares in the input image. The input image will be learnable through some uses of Neuron. This will create a feature map or activation map. The output image and function map will be input as input Data about the next layer of convolution.

The pooling layer reduces the dimension of each activation card, but still contains the most important information. The input image is divided into a set of non-overlapping rectangles. Each range is downsampled by non-linear operations such as mean and max. This layer provides better generalization, faster convergence, is robust against transformations and distortions, and is usually placed between convolution layers.

The term fully connected layer (FCL) refers to the fact that all filters in the previous layer are connected to all filters in the next layer. Convolution, pooling, and output from the ReLU layer are high-level features of the input.image. The purpose of using FCL is to use Functions for classifying input images into different classes Based on training dataset. FCL is considered a final pooling layer that uses Softmax® to supply characteristics to the classifier Activation function. Total output probabilities from The fully connected layer is 1. This is achieved using. certainty Softmax as an activation function. Softmax The function takes a vector of arbitrary real-valued scores and narrows it down to a vector of values ​​between 0 and 1 to sum NS.

**Sequential Model**-The sequential model is suitable for a simple layer stack with one input tensor and one output tensor for each layer.

**Convo2D**-The required Conv2D parameter is the number of filters that the convolution layer learns. This is an integer value and also determines the number of convolution output filters.

**Padding**- Padding parameters in the Keras Conv2D class can have either `valid` or ' same'. Setting the value to the parameter "valid" means that the input volume is not filled with zeros and the convolution can be used naturally to reduce the spatial dimension. In this case we have used it as ‘same’.

**RELU activation function**- The rectified linear activation function (ReLU for short) is a piecewise linear function that outputs the input directly if the input is positive, otherwise it outputs zero. Models that use it are easy to train and often improve performance, making them the standard activation function for many types of neural networks.

Due to the vanishing gradient problem, sigmoid and hyperbolic tangent activation functions are not available in networks with many layers. The rectified linear activation function overcomes the vanishing gradient problem, allowing the model to learn faster and perform better.The rectified linear activation is the standard activation in the development of multi-layer perceptrons and convolutional neural networks.

**RELU- g(z) = max{0, z}**

**Max-pooling layer**- Max pooling operation operation for 2D spatial data. Reduce the input along its spatial dimensions (height and width) by using the maximum value through the input window (size defined by pool\_size) for each channel of input. The step moves the window along each dimension.

**Dropout layer-** Dropout is a technique used to prevent overfitting of models. The dropout works by randomly setting the output edge of the hidden unit (the neurons that form the hidden layer) to 0 each time the training phase is updated.



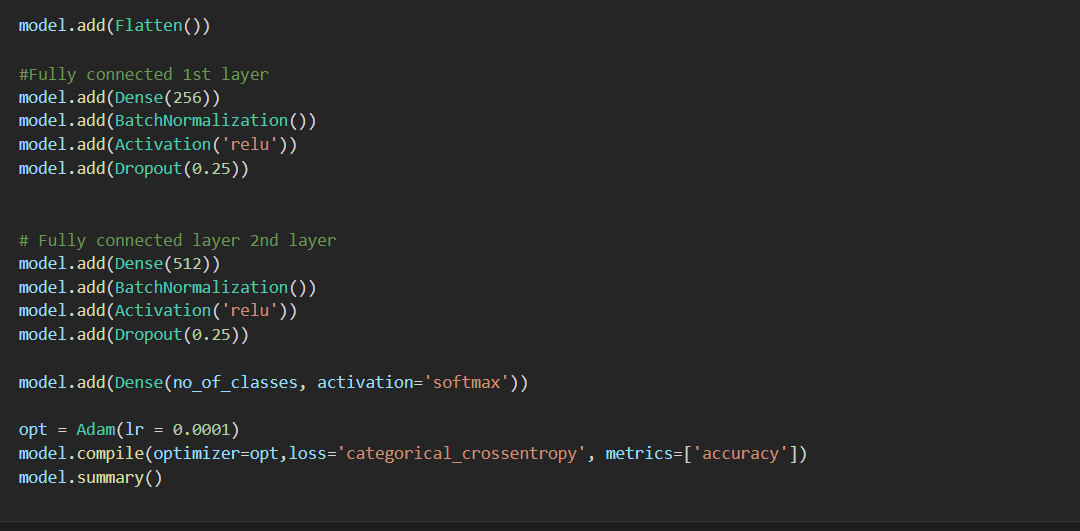
**Dense Layer**- The dense layer is a deeply connected neural network layer. That is, each neuron in the dense layer receives input from all neurons in the previous layer. High density layers are the most commonly used layers in your model.

In the background, dense layers perform matrix vector multiplication. The values ​​used in the matrix are actually parameters that can be trained and updated with the help of backpropagation.

The output produced by the dense layers is a `m` dimensional vector. Therefore, dense layers are basically used to change the dimensions of the vector. Dense layers also apply operations such as rotation, scaling, and translation to the vector.

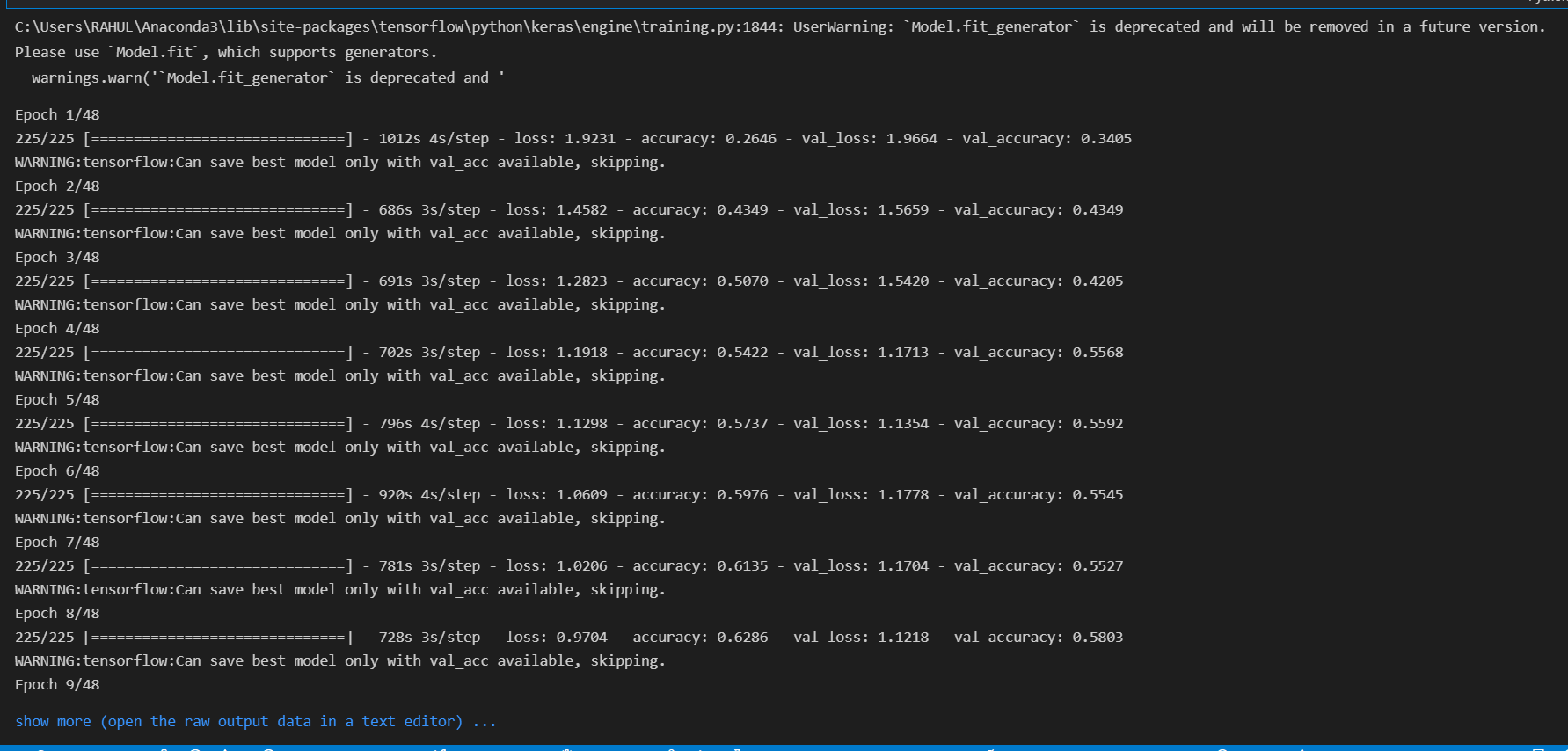
**Flatten Layer-** Flattening converts the data into a one-dimensional array and inputs it to the next layer. Flatten the output of the convolution layer to create a single long feature vector. And it is tied to a final classification model called a fully connected layer.

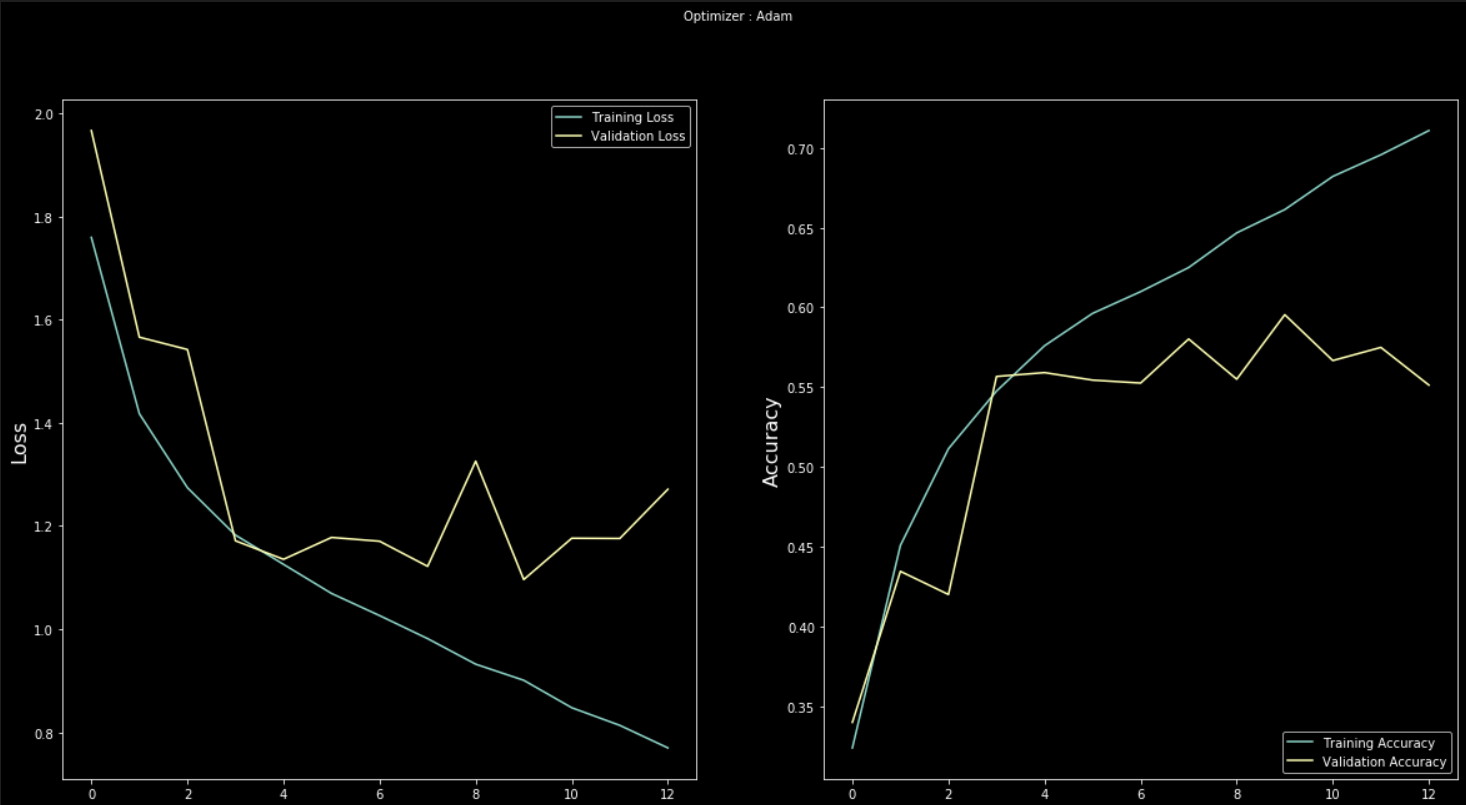
**categorical\_crossentropy loss function**- Used as a loss function for multi-class classification models with more than one output label. One-hot encoding values ​​in the form 0 and 1 are assigned to output labels. Output labels are converted to categorical coding using Keras, if available in integer format.



**Adam Optimizer**- Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first order and second order moments.

VI. CODE IMPLEMENTATION FOR MODEL





**Saving the file as model.h5:**

model\_json = model.to\_json()

with open("model.json", "w") as json\_file:

json\_file.write(model\_json)

model.save\_weights("model.h5")

print("Saved model to disk")

from tensorflow.keras.models import load\_model

model.save("model.h5")

VII. SOFTWARE REQUIREMENTS

* For website development, we will be needing python and flask libraries.

* Additionally, we will be needing libraries like TensorFlow and Keras for developing the training and testing model.
* To execute we will be requiring visual studio code.

VIII. EMOTIONS USED

Happy: This emotion determines the happy behaviour of the user.

Sad: This emotion determines the sad behaviour of the user.

Neutral: This emotion determines the neutral behaviour of the user.

Anger: This emotion determines the anger behaviour of the user.

Disgust: This emotion determines the disgust behaviour of the user.

Fear: This emotion determines the fear behaviour of the user.

Surprise: This emotion determines the surprise behaviour of the user.

IX.CODE IMPLEMENTATION FOR

EMOTION RECOGNITION

def start\_app(frame):

font = cv2.FONT\_HERSHEY\_SIMPLEX

facec = cv2.CascadeClassifier('./haarcascade\_frontalface\_default.xml')

cnn = FacialExpressionModel("model1.json", "chkPt1.hdf5")

#while True:

#faces, fr, gray\_fr = \_\_get\_data\_\_(frame)

gray = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

faces = facec.detectMultiScale(gray, 1.3, 5)

for (x, y, w, h) in faces:

fc = gray[y:y+h, x:x+w]

fc = cv2.normalize(fc,None,0,255,cv2.NORM\_MINMAX)

# fc = cv2.addWeighted(fc,1.5,blur,-0.5,0)

roi = cv2.resize(fc, (48, 48))

pred, lbl = cnn.predict\_emotion(roi[np.newaxis, :, :, np.newaxis])

#print(pred)

#list.append(pred)

cv2.putText(frame, pred, (x, y), font, 1, (255, 255, 0), 2)

cv2.rectangle(frame,(x,y),(x+w,y+h),(255,0,0),2)

x1 = x + w//2

y1 = y + h//2

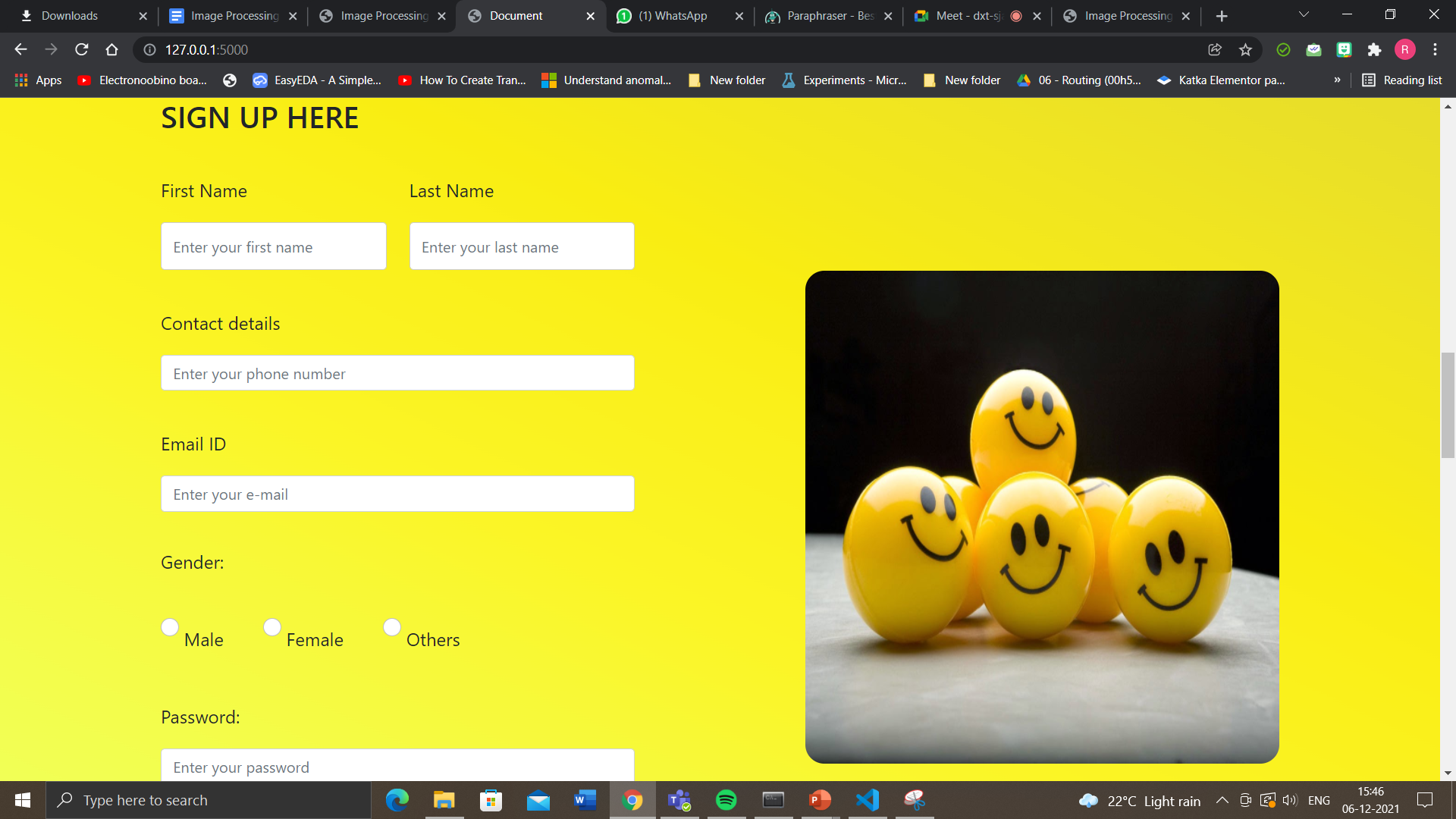
#emo = emoji[lbl]

return frame,pred

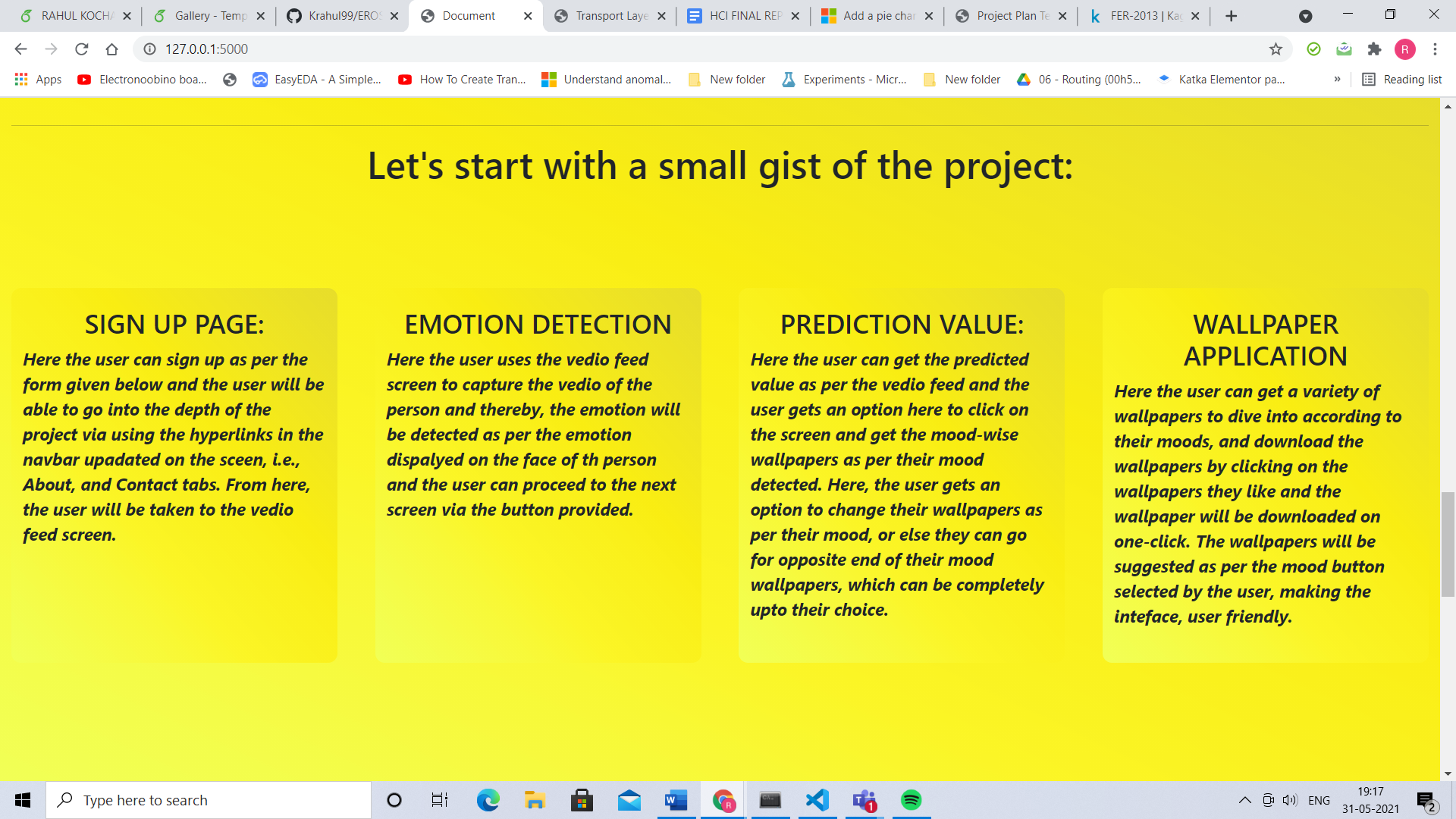
X.RESULT



This section describes the cover page of our web-app.



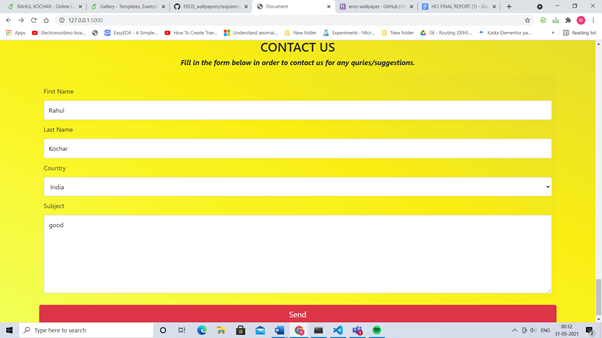
We have a section on the first page where the user can get the information about how our website works.



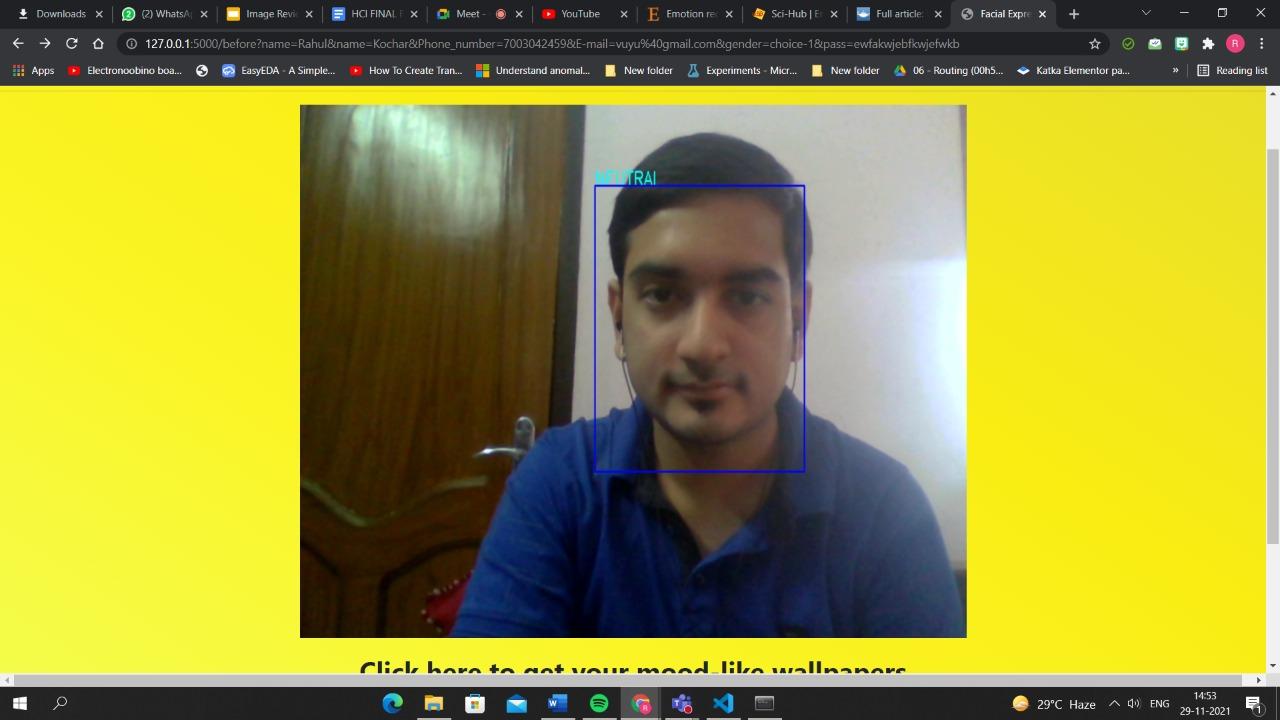
We have a section on the first page where the user can get the information about how our website works

Then we have a contact us form so that the user after using the website can give us feedback or any kind of support or help can be also taken care of. The predicted value of our emotion is being shown (here neutral) as well as when the user clicks on the given emotion emoji the user gets into the desired wallpaper store. We enter into the third page where the predicted value of our emotion is being shown (here neutral) as well as when the user clicks on the given emotion emoji the user gets into the desired wallpaper store.

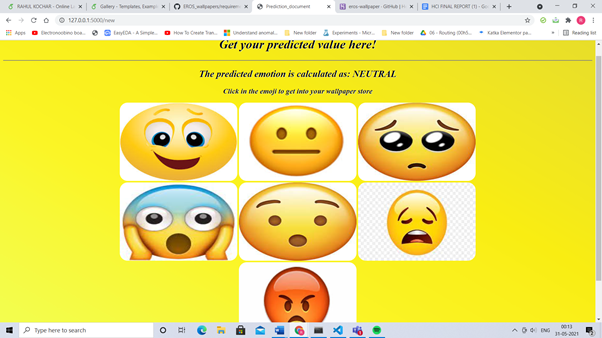
The wallpaper store of neutral emotion. Here the user can download its favorite wallpaper and use them wherever they want. A log-out button is also provided to get the user back to the sign in page.



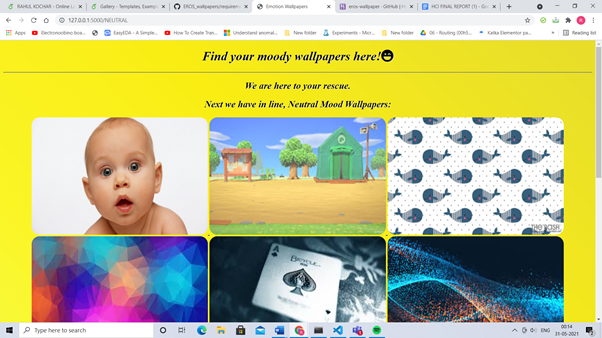
The predicted value of our emotion is being shown (here neutral) as well as when the user clicks on the given emotion emoji the user gets into the desired wallpaper store.



We enter into the third page where the predicted value of our emotion is being shown (here neutral) as well as when the user clicks on the given emotion emoji the user gets into the desired wallpaper store.



The wallpaper store of neutral emotion. Here the user can download its favorite wallpaper and use them wherever they want.



A log-out button is also provided to get the user back to the sign in page.

XI. CONCLUSION

Facial expression recognition is a challenging problem in the field of image analysis and computer vision that has received a great deal of attention over the last few years because of its many applications in various domains.

This proposes a human facial expression recognition model based on the eigenface approach in which the various emotions are recognized by calculating the Euclidean distance between the input test image and the mean of the eigenfaces of the training dataset.

The field of research in expression recognition is an area which can be further explored and improved. This project focuses on directly transforming the dataset images to their eigen faces so as to depict a general sense in which these expressions are formed.

Finally, after interacting with the user and then predicting the emotion of the user, our project is capable of suggesting the relative wallpapers based on the emotion/expression detected and allows the user to download it too.

1. *References*

[1] <https://sci-hub.se/10.1109/MEES.2017.8248937>

[2]<https://sci-hub.se/https://doi.org/10.1145/1027933.1027968>

[3][https://sci-hub.se/https://doi.org/10.1016/j.procs.2017.05.02](https://sci-hub.se/https://doi.org/10.1016/j.procs.2017.05.025)

[4] <https://sci-hub.se/10.1109/ICICS.1997.647126>

[5]<https://sci-hub.se/10.1109/ICECA49313.2020.9297483>

[6][https://link.springer.com/chapter/10.1007/978-3-319-04960-1](https://link.springer.com/chapter/10.1007/978-3-319-04960-1_2)