MINOR REPORT: EVEN SEMESTER 2020

EMOTION BASED STRESS ANALYZER

FACULTY SUPERVISOR : MS.NEHA BANSAL

SUBMITTED BY:

DEEKSHA VARMA (17103018)

PRAGYA SHARMA (17103022)

KHUSHI UTKARSH (17103124)

SUBMITTED TO:

TABLE OF CONTENTS

DECLARATION

We, the undersigned, hereby declare that the project titled **"Emotion Based Stress Analyzer"** and its associated report written and submitted by us to Jaypee Institute of Information Technology, Noida, in partial fulfillment of the requirement for the award of degree of Bachelors of Technology under the guidance of Ms. Neha Bansal is our original work and the conclusions drawn therein are based on the material collected by ourselves.

We further declare that the work reported in this project has not and will not be submitted, either in part or in entirety, for the award of any other degree or diploma in this institute or in any other institute or university.

(Deeksha Varma) (Khushi Utkarsh) (Pragya Sharma)

Date : 14 May, 2020

ACKNOWLEDGEMENTS

Indebtedness should feel like a dead weight, especially when associated with something that one does to examine their worth, and yet, thanks to everyone who associated themselves with this project and with us, it feels liberating. Ergo, we would like to take a few moments to express our immense gratitude to God and the following people for their gratuitous aid and constant support.

First and foremost, we would like to thank our mentor **Ms. Neha Bansal and Dr. Ankit Vidyarthi,** for being ever so patient and educative as they guided us along the path of conceptualising an abstract idea and converting it into a real, working model. Thank you ma'am, for your kindness and tenacity with us, considering it could not have been easy working with three amateurs and yet, you managed to do so fluidly as one could possibly imagine. It was a real honour to have been mentored by you.

We would also like to extend a token of gratitude to **Dr. Dharamveer Rajpoot** and **Ms. Mradula**, both of whom constituted the panel for the mid-evaluation of our project. Thank you ma'ams, for providing us with your priceless insights and giving us a sense of direction and convergence.

Special thanks to the panel for being empathetic, candid and considerate with us. We hope we are able to stand by your expectations with this project, if not exceed them.

A huge shoutout to our parents, who provided us with all the necessary resources required to finish this project. We would also like to express our gratitude to Jaypee Institute of Information Technology for having provided us with the opportunity of working on this project.

Lastly, to anyone who supported us in small ways, directly or indirectly related to this project: Thank you. We couldn't have done it without you, any of you.

INTRODUCTION

Stress is a term frequently utilized synonymously with negative life experiences or life occasions. Logical research on pressure and uneasiness offers different points of view on the issue. The expanding pace of life hurried and focused ways of life imply that stress is an integral part of human life. A man in a condition of adjusting to pressure demonstrates conduct resistances. This prompts changes in one's psychological procedures and enthusiastic scene. Stress can be a kind of mental ordeal. It additionally has a role in response to nature, and inspirations. A large damage may be there due to measures of pressure.

Due to stress there may be other health issues like obesity, heart attack, diabetes, asthma etc. Every hour, a student commits suicide in the different part of the country. Our country ranks on 19th position in the list of sucide rates according to the WHO report of 2016 . This is an approach with the help of which we can analyze the stress at its very first step. If we can find out the stress level in the students, in the short or long term we can help them in recovering.
Our inspiration for this project is the expansion in the number of suicide in our nation.

Within the past decade, there has been significant effort toward automatic recognition of human facial expression using computer vision. This facial expression detection has mostly been used for emotion detection alone. We have aimed to take it a step forward, we will try to detect if a person is stressed or not. And if he is, we will detect to what extent he is stressed.

Our project will detect the stress level based on what emotion he is experiencing.

The idea behind this project is that it may be used in the following areas:

1. Used in basic stress detection:-
   - For students in schools as well as universities
   - For patients in hospitals
   - For employees in an organisation
   - For inmates in jail

   All the above information can tell how stressed a person is in a social setting. This information can be used by the authorities for the betterment of their services and for the contentment of people.

2. Used in Medical Examination:-
   In the psychiatric domain of medicine there are many diseases which involve hyped emotions. One example can be Bipolar disorder. The bipolar disorder includes sudden hyped emotions which can be sensed using our emotion detector.

3. In the quarantine period:-
   Nowadays even the important things like recruitment processes or business meetings and even daily college and school classes are shifting to video conferencing. When you meet someone in person then you are better able to gauge the emotion of a person. But now that video conferencing has become such a major part of our daily life a technology which can tell if a person is stressed would be very helpful.

PROBLEM STATEMENT

Suicides are an impulsive reaction due to stress, for example, there may be money related challenges, several issues with one's connection or partner, or maybe due to bullying. Once a person thinks about it several different methods come into his mind, for example, guns, medications, and toxin.
In India, about 46,000 suicides occurred each in 15–29 and 30–44 age groups in 2012 – or about 34% each of all suicides.
Increasing rate of suicide is becoming an area of concern for the government and also the people of the nation. Treatment of mental clutters is one of the ways to deal with diminish the rate of suicide in the coming year.
The **aim of this project** is to capture human emotion in real time to detect stress which can help the concerned authorities to take quicker steps to improve mental health.

RESOURCED USED

1) OpenCV :

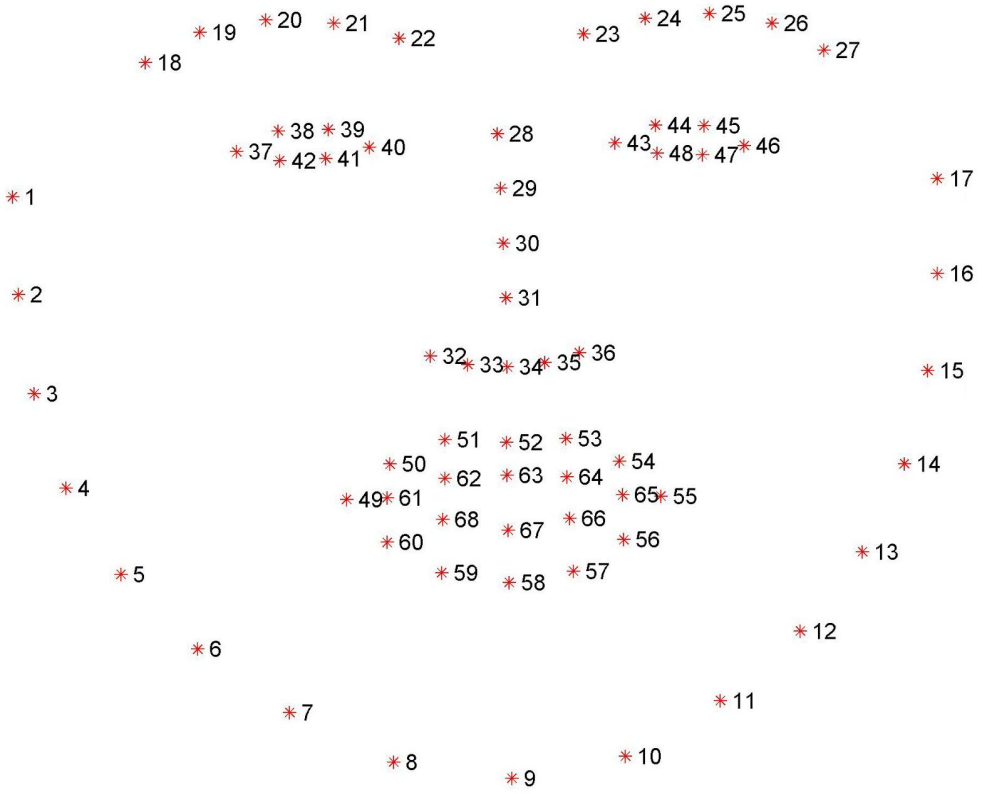
OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state- of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in video etc.

2) Dlib :

Dlib is a modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems. It is used in both industry and academia in a wide range of domains including robotics, embedded devices, mobile phones, and large high performance computing environments. Dlib's open source licensing allows you to use it in any application, free of charge.

dlib library helps us for face detection as it has a detailed facial landmark

The facial landmarks produced by dlib follow an indexable list :



We can therefore determine the starting and ending array slice index values for extracting (x, y)-coordinates for both the left and right eye

3) Imutils:

A series of convenience functions to make basic image processing operations such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV and Python.

4) Dataset:

One can download the facial expression recognition (FER) data-set from Kaggle challenge here. The data consists of 48×48 pixel gray scale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

The training set consists of 35,888 examples. train.csv contains two columns, “emotion” and “pixels”. The “emotion” column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The “pixels” column contains a string surrounded in quotes

for each image. The contents of this string a space-separated pixel values in row major order

5) Loading FER Data-set

The below code loads the data-set and pre-process the images for feeding it to CNN model. There are two definitions in the code snippet here:

1. **def load_fer2013** : It reads the csv file and convert pixel sequence of each row in image of dimension 48*48. It returns faces and emotion labels.

2. **def preprocess_input**: It is a standard way to pre-process images by scaling them between -1 to 1. Images are scaled to [0,1] by dividing it by 255. Further, subtraction by 0.5 and multiplication by 2 changes the range to [-1,1]. [-1,1] has been found to be a better range for neural network models in computer vision problems.

6) Keras.preprocessing.image

The images collected from the live cam have to be processed. We need to convert them into numpy arrays so that they can be used to predict the emotions and further analysis. In Keras this can be done via the Keras.preprocessing.image class.

7) Keras.models

Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow. It was developed to make implementing deep learning models as fast and easy as possible for research and development.

8) Convolutional neural network

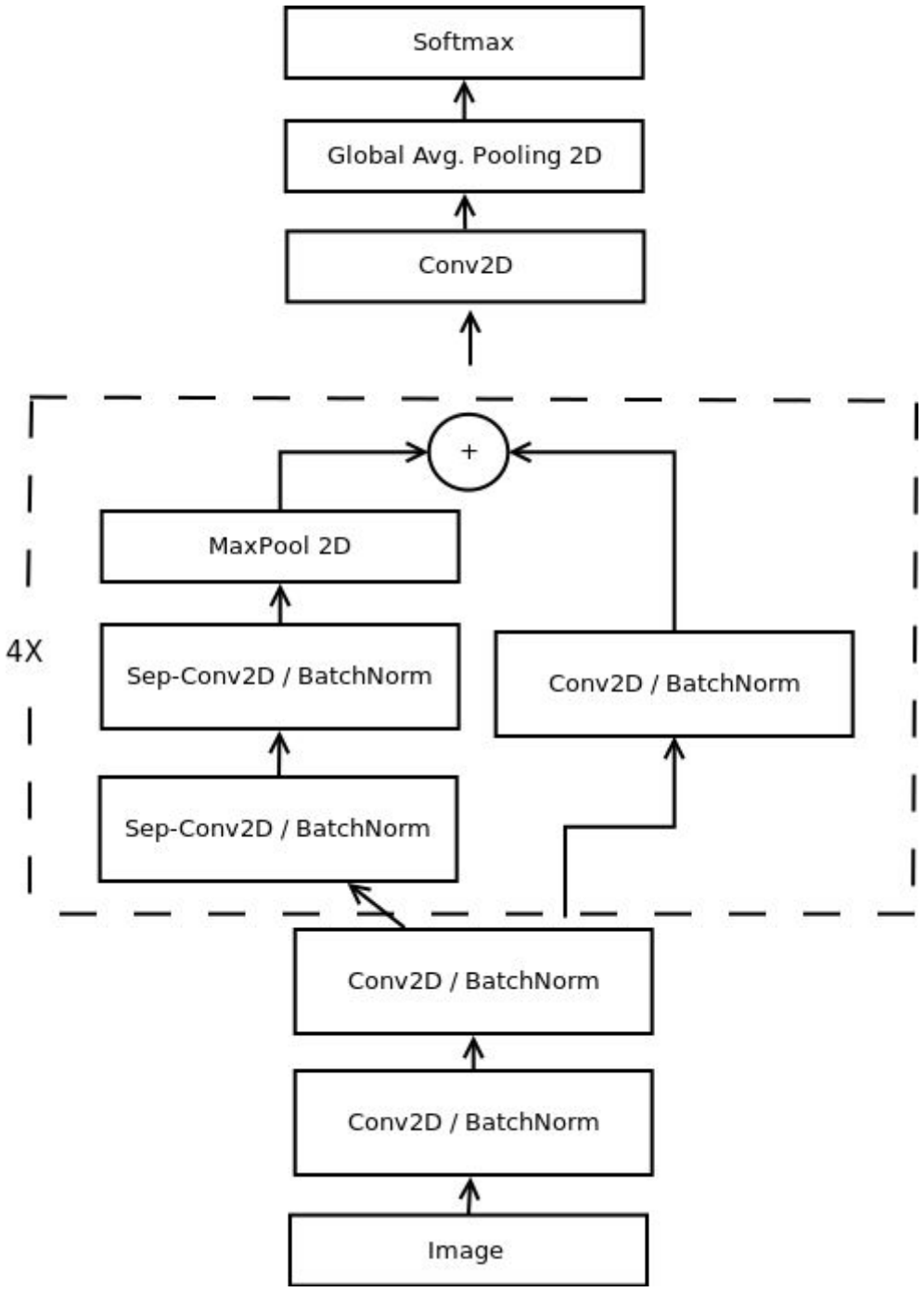
In neural networks, Convolutional neural networks (ConvNets or CNNs) is one of the main categories to do image recognition, images classifications. Objects detections, recognition faces

etc., are some of the areas where CNNs are widely used.

CNN image classifications take an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers see an input image as an array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.

9) Mini Xception Model

Here comes the exciting architecture which is comparatively small and achieves almost state-of-art performance of classifying emotion on this data-set. The below architecture was proposed by Octavio Arragia et al. in this paper.



Proposed Mini_Xception architecture for emotion classification

One can notice that the center block is repeated 4 times in the design. This architecture is different from the most common CNN architecture like one used in the blog-post here. Common architectures use fully connected layers at the end where most of the parameters reside. Also, they use standard convolutions. Modern CNN architectures such as Xception leverage from the combination of two of the most successful experimental assumptions in CNNs: the use of residual

modules and depth-wise separable convolutions.

There are various techniques that can be kept in mind while building a deep neural network and are applicable in most of the computer vision problems. Below are few of those techniques which are used while training the CNN model below.

1. Data Augmentation : More data is generated using the training set by applying transformations. It is required if the training set is not sufficient enough to learn representation. The image data is generated by transforming the actual training images by rotation, crop, shifts, shear, zoom, flip, reflection, normalization etc.
2. Kernel_regularizer : It allows you to apply penalties on layer parameters during optimization. These penalties are incorporated in the loss function that the network optimizes. Argument in the convolution layer is nothing but L2 regularisation of the weights. This penalizes peaky weights and makes sure that all the inputs are considered.
3. BatchNormalization : It normalizes the activation of the previous layer at each batch, i.e. applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. It addresses the problem of internal covariate shift. It also acts as a regularizer, in some cases eliminating the need for Dropout. It helps in speeding up the training process.
4. Global Average Pooling : It reduces each feature map into a scalar value by taking the average over all elements in the feature map. The average operation forces the network to extract global features from the input image.
5. Depthwise Separable Convolution : These convolutions are composed of two different layers: depth-wise convolutions and pointwise convolutions. Depth-wise separable convolutions reduces the computation with respect to the standard convolutions by reducing the number of parameters. A very nice and visual explanation of the difference between standard and depth-wise separable convolution is given in the paper.

LITERATURE SURVEY

Research papers on Facial Recognition				
Title	Technique	Dataset	Performance	Remarks
Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks[1]	Feedforward Deep Convolution Neural Networks (FDCNN)	-Emotion Dataset This dataset containing five different emotions (happy, angry, sad, untrustworthy and fear) and actions (jumping, sitting, walking) performed by 29 actors -GEMEP Dataset The Geneva Multimodal Emotion Portrayals (GEMEP) is a set of five basic emotions (Angry, Joy, Fear, Sad and Pride) expressed by 10 actors have been used for this work	95.4% is the accuracy	This model is better than the baseline models. This model is representing deep convolutional features to extract saliency information at multiple scales
Automatic Facial Expression Recognition Using DCNN[2]	Deep Convolution Neural Networks (DCNN)	All 213 facial expressions images that are available with the JAFFE dataset	98.12% which is very high	This is the highest accuracy that we have observed so far. This model works for both static as well as video dataset.
Statistical Moments based Facial expression Analysis[3]	Feature Extraction: Zernike moments Classification: Naive Bayesian classifier	JAFFE (Japanese Female Facial expression) database 60 images used for experiment	Average accuracy for six emotions is 81.66% in time less than 2 seconds	Emotion accuracy graph shows highest recognition rate of happiness and lowest recognition rate of sadness
Facial expression recognition with Auto-Illumination correction [4]	Expressions on the face are determined with Action Units (AU's)	Single and Multiple face image	60% recognition rate for multiple face image	Illumination on image plays vital role.

Identification driven Emotion recognition system for a Social Robot[5]	Hybrid approach used for personalized emotion recognition,	MUG facial expression database used. More than 50 people in the frontal face database used aged between 20-25 years.	82% performance achieved with KNN Classifiers.	3D model facial image used. KNN classifier gives good performance for emotion recognition
The application study of learner's face detection and location in the teaching network system based on emotion[6]	SVM(Support Vector Machine) classifier based Adaboost algorithm used	PIE face image database used	Detection and Correction rate 95% or more.	Presents application of face emotion recognition of the E-learning system.
Cognitive Face Analysis System for Future Interactive TV[7]	Ada-LDA learning algorithm and MspLBP features used for effective multi-class pattern classifier	JAFFE and MIT+CMU database	Recognition rate of over 15 frames per second	Real time performance with high recognition rate

An Efficient Algorithm for Motion Detection Based Facial Expression Recognition using Optical Flow[8]	Infra-Red (IR) illumination used for facial features approximately localization. Source Vector (SV) used for vector collection and identification of emotion is based on highest degree of similarity between source vector and execution motion vector	Approximately 1000 images sequences of Cohn-Kanade Facial Expression Database with 65% female facial image used for experiment	94% recognition rate	Only three frames are sufficient to detect facial expression.
Mental Stress Detection in University Students using Machine Learning Algorithms[9]	Used classifiers like Naive Bayes, Random Forest,SVM, K-NN.	The dataset was taken from the 206 students of Jaypee Institute of Information Technology Noida.We have classified the data in two conditions one is before the exams and other is stress due to the usage of the internet. The dataset was collected for a PSS test which includes 14 questions overall including the entire emotional question. The marking for the questions was in 5 ways (a) Never (b) Almost Never (c) Sometimes (d) fairly Often (e) Very Often.	Random Forest-83.33, Naive Bayes-71.42, SVM-87.71, KNN-55.55.	The accuracy is good except for the KNN.

Development of a Stress Classification Model Using Deep Belief Networks for Stress Monitoring[10]	DBN Model,Naive bayes, Decision tree, SVM	KNHANES VI dataset,	DBN-66.23 NB-66.23 DT-66.28 SVM-66.02	The accuracy of the DBN model is similar to the other models used but DBM saves time because the semi-supervised learning technology allows the use of unlabeled training samples.

IMPLEMENTATION

Step 1 – Fetching image from web camera

Using the OpenCV and Dlib in Python we are able to capture an image from our laptop webcam and by using the dlib.get_frontal_face_detector() and the dlib.shape_predictor() modules.

In this webcam consecutive 5 frames are taken into consideration for detecting one blink.

The dlib.get_frontal_face_detector() fetches the frontal face detector.

While the dlib.shape_predictor() fetches the prediction function from the Dlib library.

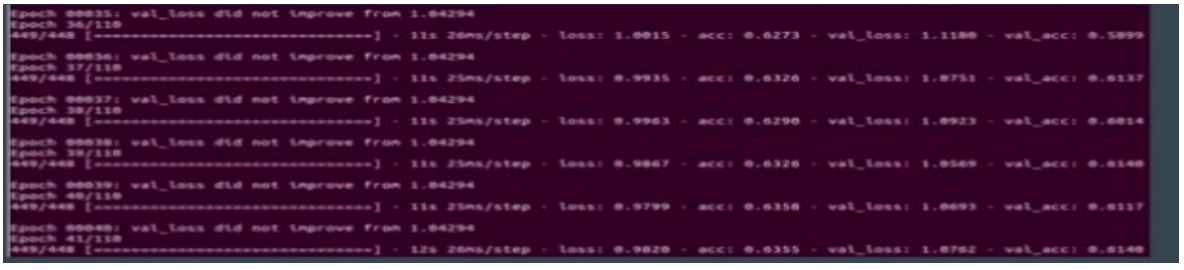
Step 2 - Image preprocessing

The image captured by the webcam has to be preprocessed before using it for emotion recognition and stress detection. There are 5 consecutive frames that are selected from the camera. The image has to be transformed in such a way that it is readable for the machine. The frame collected from the camera is firstly converted into a grayscale image. A grayscale image is always better for the preprocessing. After converting the color the image is flipped and then resized according to the data set.

Step 3 - Emotion Recognition

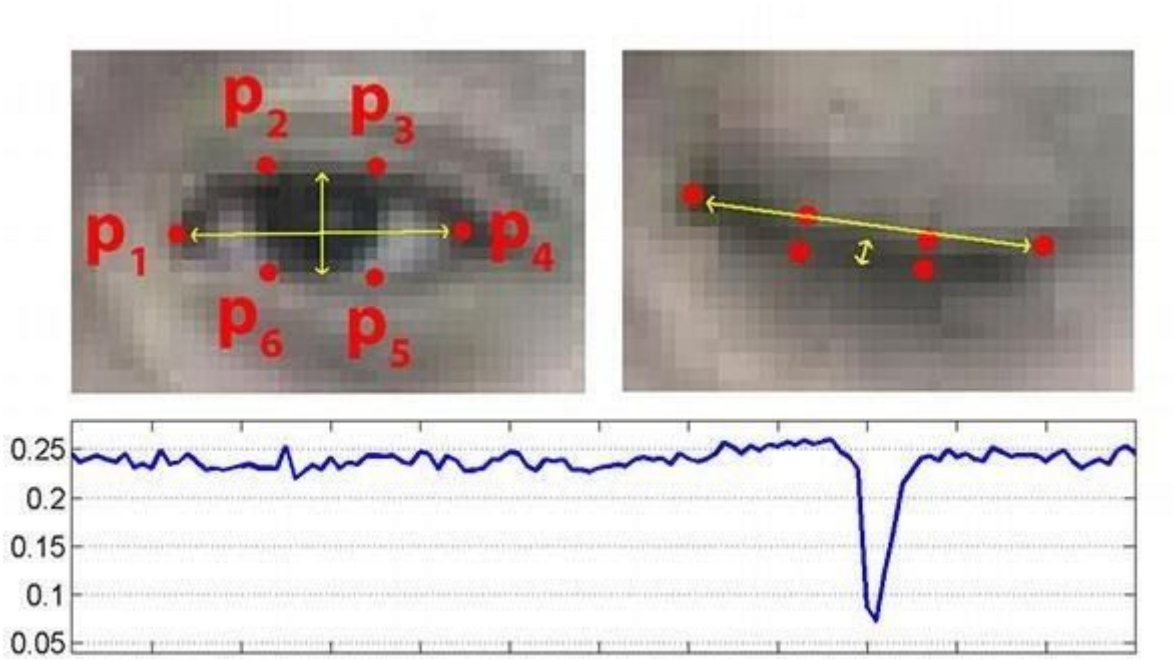
Humans are used to taking in non verbal cues from facial emotions. Now computers are also getting better at reading emotions. So how do we detect emotions in an image? We have used an open source data set — Face Emotion Recognition (FER) from Kaggle and built a CNN to detect emotions. The emotions can be classified into 7 classes — happy, sad, scared, disgust, angry, neutral and surprised.

If one of these emotions are being seen as sad or angry then we go for further testing of the image to see if the person is stressed or not. Otherwise if any other emotion is seen i.e happy,fear,disgust,surprise,neutral then the results will show not stressed.



Step 4 – Eyeblink detection

Using the OpenCV and Dlib library in Python we detect the eye coordinates from the face_utils.FACIAL_LANDMARKS_IDXS function then extract the image with reference to the coordinates from the frame. Now this function keeps a count of the number of times the eyelids were shut.



Plotting the eye aspect ratio over time. The dip in the eye aspect ratio indicates a blink

we can then derive an equation that reflects this relation called the eye aspect ratio (EAR):

$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

On the top-left we have an eye that is fully open — the eye aspect ratio here would be large(r) and relatively constant over time.

However, once the person blinks (top-right) the eye aspect ratio decreases dramatically, approaching zero.

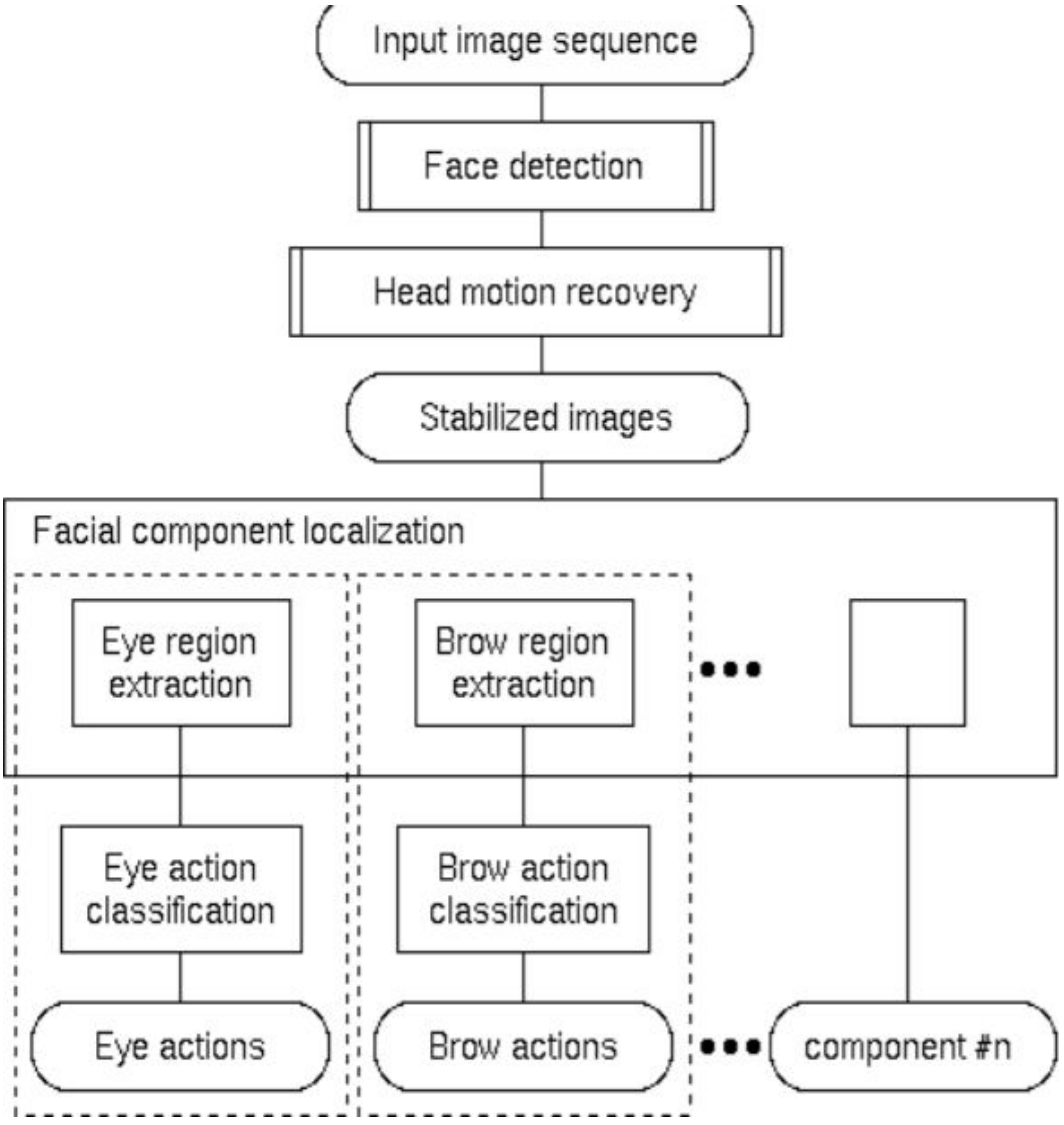
The bottom figure plots a graph of the eye aspect ratio over time for a video clip. As we can see, the eye aspect ratio is constant, then rapidly drops close to zero, then increases again, indicating a single blink has taken place.

In a study it was seen that a person's normal blinking capability is 8-21 times in a minute. If the frequency is more than this then the person can be stressed. This doesn't happen all the time but in maximum cases this is the norm. Considering it to be a general case if the count of the eye blink is more than 25 we categorise the person as being stressed.

Step 4 – Eyebrow detection

Using the OpenCV and Dlib library in Python we detect the Eyebrow coordinates from the face_utils.FACIAL_LANDMARKS_IDXS function then extract the image with reference to the coordinates from the frame. Now we have the eyebrows extracted using cv2.convexHull and cv2.drawContours. Now we have calculated the euclidean distance using dist.euclidean(coordinates of each eyebrow) of eyebrows in each frame and then on the basis of the previous results each frame value has been normalised. The stress value is then calculated using the distance. If the value is >=80 then it shows stressed otherwise not stressed.

IMPLEMENTATION IMAGE:



RESULT AND ANALYSIS

a) Eye Blink Detection- This has 2 output

- Displays the number of blinks,the EAR (Eye Aspect Ratio) and the state.

1)

Number of blinks:09
EAR:0.38
Not Stressed

2)

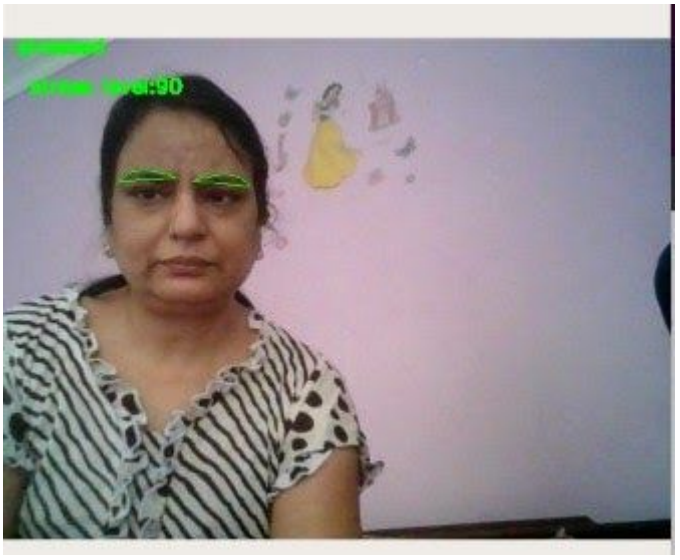
Blinks:23
EAR:0.37
Stressed

b) Eyebrow Detection- This has 2 outputs

1) Not Stressed with the stress level as shown the picture

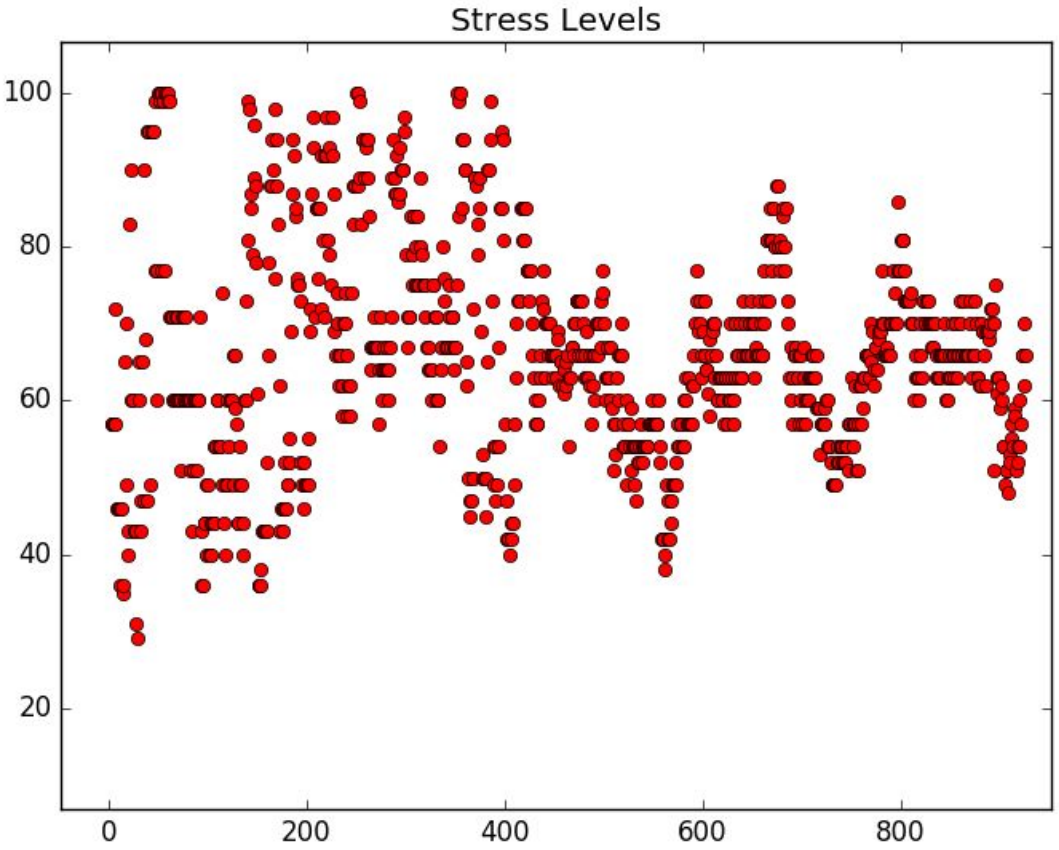
Not Stressed
Stress level:76

2) Stressed with the stress level as shown in the picture



Stressed
Stress level:90

VISUAL REPRESENTATION OF STRESS LEVELS WITH RESPECT TO TIME FRAME



On the x-axis we have the time frames.
On the y-axis we have the different levels of stress detected throughout the execution frame.

This helps us to analyse at what particular instance the person was highly stressed and under pressure.
It also shows to what extent the subject was stressed.

CONCLUSION

Nowadays people's mental health is deteriorating at an alarming rate which hampers with their well being. Our project aims to help in detection of stress in daily life and help organisations, schools, and hospitals improve their working environment.

Our proposed project has been systematically built in order to increase the efficiency of determining the stress level. We have developed a vision system that performs face detection,emotion classification and Stress level in a single integrated module.Each module working independent of the other thus increasing the reliability of the result. Our complete pipeline has been successfully integrated by the mini xception model of CNN. We can find the stress level and with its help one can perform initial analysis to help the person in his/her initial stages of stress if the person is in a high mental stress state.

The efficiency of our project lies between 65-70% which can be improved if we add more facial features for example head orientation, gaze movement, lip and hand movement.

FUTURE SCOPE

1. The orientation of the face is also an important factor as to how a person

is feeling. If a person is embarrassed he tends to bow his head. If a person is thinking then he tends to tilt his head upwards to one side. If a person is proud then he tends to lift his head up high. Face orientation can tell alot about how a person is feeling adding this to our project would enhance its working and improve the efficiency of the already existing code.

2. Many people have a habit of hand movement while talking. These gestures can also tell alot about how a person is feeling. For example when a person is happy then he tends to move his hands with energy and excitement. When a person is nervous he tends to fidget with his fingers. When a person is frustrated he tends to move his hands with vigour. This movement of muscles can help us improve the efficiency even more.The hand movement will also help us determine the body language.

3. On the face after eyes lips are the most expressive feature. When we are happy lips spread into a smile. When one cries the lips are pressed downwards.And when a person is stressed the lips are pursed. Therefore, lips can also tell us a lot about human emotion. Our final feature would be lip placement reading.

REFERENCES

[1] https://www.sciencedirect.com/science/article/pii/S1877050919306908

[2] https://www.sciencedirect.com/science/article/pii/S1877050916314752

[3] https://ieeexplore.ieee.org/document/7154768

[4]https://www.researchgate.net/publication/271425201_Facial_Expression_Recognition_with_Auto-Illumination_Correction

[5]https://www.researchgate.net/publication/261455573_Identification-driven_emotion_recognition_system_for_a_social_robot

[6] https://ieeexplore.ieee.org/document/5480962

[7] https://ieeexplore.ieee.org/document/5373798

[8]https://www.researchgate.net/publication/242597024_An_Efficient_Algorithm_for_Motion_Detection_Based_Facial_Expression_Recognition_using_Optical_Flow

[9]https://www.sciencedirect.com/science/article/pii/S1877050919306581

[10] https://www.researchgate.net/publication/321055983_Development_of_a_Stress_Classification_Model_Using_Deep_Belief_Networks_for_Stress_Monitoring