

Detection of Stress Using Image Processing and Machine Learning Techniques

Nisha Raichur^{#1}, Nidhi Lonakadi^{*2}, Priyanka Mural^{#3}

Department of Information Science and Engineering, BVBCET, Hubli, India

*Tata Elxsi, ITPL main road, Hoodi Bengaluru, India

¹nishalraichur@gmail.com

* Accenture, Ecospace Bellandur Bengaluru, India

²nidhilonakade@gmail.com

* Infosys, Pocharam Village Hyderabad India

³priyankamural76@gmail.com

Abstract— Stress is a part of life it is an unpleasant state of emotional arousal that people experience in situations like working for long hours in front of computer. Computers have become a way of life, much life is spent on the computers and hence we are therefore more affected by the ups and downs that they cause us. One cannot just completely avoid their work on computers but one can at least control his/her usage when being alarmed about him being stressed at certain point of time. Monitoring the emotional status of a person who is working in front of a computer for longer duration is crucial for the safety of a person. In this work a real-time non-intrusive videos are captured, which detects the emotional status of a person by analysing the facial expression. We detect an individual emotion in each video frame and the decision on the stress level is made in sequential hours of the video captured. We employ a technique that allows us to train a model and analyze differences in predicting the features. Theano is a python framework which aims at improving both the execution time and development time of the linear regression model which is used here as a deep learning algorithm. The experimental results show that the developed system is well on data with the generic model of all ages.

Keyword - Stress, Facial expression, Theano, Framework, Deep learning

I. INTRODUCTION

Most of the researchers focused on detecting stress involved in a person, which causes in a person several emotional problems like anxiety, grief, low self-esteem and other mental health problems. Recent studies have shown that stress can also affect the aspects of your life, including your thinking ability and physical health. To reduce riskiness from being stress and affected with its adverse effects, it is crucial to detect such emotions and take certain actions to relax them.

Most of the previous work on stress detection is based on the digital signal processing, taking into consideration Galvanic skin response, blood volume, pupil dilation and skin temperature. And the other work on this issue is based on several physiological signals and visual features (eye closure, head movement) to monitor the stress in a person while he is working. However these measurements are intrusive and are less comfortable in real application.

In this work we develop a stress detection system based on the analysis of the facial expression. The system is non-intrusive and is able to run in real-time. A camera is used to capture the near frontal view of the person while he is working in front of the computer. The camera is mounted facing a person. Video captured is divided into three sections of equal length and set of equal number of image frames are extracted from each section correspondingly and are analysed. The image analysis includes the calculation of the variation in the position of the eyebrow from its mean position. The displacement of eyebrow from its position is calculated by scanning the image for the eyebrow co-ordinates. If the person is found stressed in the consecutive sections of the time intervals which was previously divided, the decision for stress detection is made for a person working in front of the computer

With the obtained results we employ the technique of deep learning which is a branch of machine learning which gives the computer an ability to learn without being explicitly programmed. Theano is a python framework which aims at improving both the execution time and development time of the linear regression model which is used here as a deep learning algorithm.

II. SYSTEM OVERVIEW:

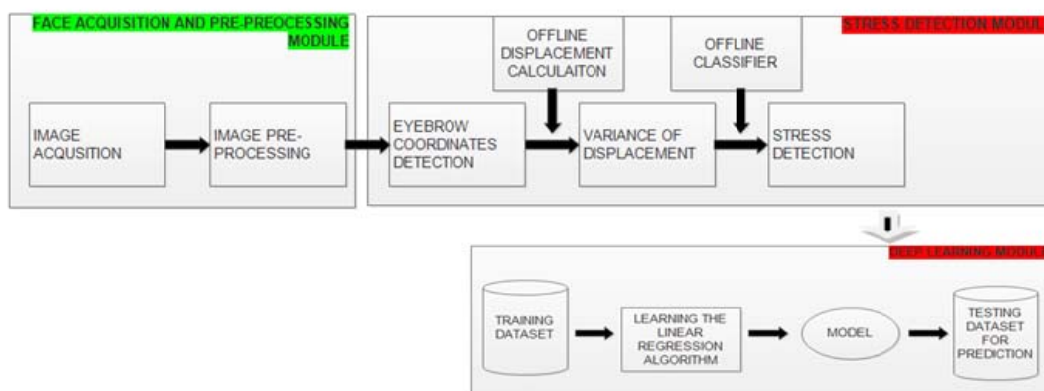


Fig. 1. Module explaining the system overview

In this section we briefly describe the hardware and software of the developed section. The camera is mounted that captures the near frontal view of a person while working in front of the computer. Camera captures the images that are compensated for the illumination effects caused by the ambient lighting conditions

The acquired video frames are processed sequentially by the software modules as illustrated in Fig 1. The face acquisition and pre-processing module pre-processes the real time captured images, which includes adjusting the brightness and contrast of the images captured which are further converted to binary images for the analysis in the subsequent module. The stress detection module scans the binary image from the extreme left top to record the co-ordinates of the eyebrow. The offline displacement calculation sub-module calculates the displacement of the eyebrow using the obtained eyebrow co-ordinates which is subsequently followed by the variance calculation of the displacement. The classifier sub-module is trained offline are employed to determine the presence of emotion. The integrated decision of individual frames eventually determines the level of stress involved.

The results that are obtained using classification are further optimized using machine learning. Deep learning is an area of machine learning that attempt to model high-level abstractions in data. Linear regression is a supervised learning algorithm which analyzes the training data and produces the inferred function, which can be used to predict the results for the test dataset. Theano being the python package has been used in training the linear regression algorithm using training dataset that produces a model to predict the desired behaviour.

III. EMOTIONAL STRESS DETECTION:

This section describes the implementation details of the individual modules. This includes the definition of the detection target and the methodologies.

A. Task Definition:

Different people may behave or express differently under stress and it is hard to find a universal pattern to define the stress emotion. And it is not easy train a model offline that classifies and predicts if a person is stressed. To ease the problem, we have evolved with an algorithm that relates to the facial area of interest in stress recognition. Eyebrow which shows a rigid transformation is used as the facial area of interest in stress recognition. The eyebrow movement has not got any universal pattern but the variation corresponding to a particular person is used to detect the stress involved. We consider that stress is detected if a person shows high variation of the eyebrows constantly within a fixed time interval.

B. Methodology:

1) Image Pre-processing:

The face acquisition module processes the video sequences captured by the camera. The image frames are extracted and the pre-processing of the images for the subsequent analysis in the further modules is done. Pre-processing of the images includes two transformations of the extracted frame. First one being the pixel transformation and the other one is binary transformation of the modified image after pixel transformation.

Pixel transformation as an image processing function is a technique of transforming a pixel value. Each output pixel value depends on only the corresponding input pixel value, which helps alter the brightness and contrast of an input image. The two commonly used point processing are multiplication and the addition with a constant.

$G(i, j) = \alpha \cdot F(i, j) + \beta$, The parameters $\alpha > 0$ and β are often called gain and the bias parameters, these parameters are said to control the brightness and the contrast of the image respectively

$F(i, j) \rightarrow$ source image pixel, $G(i, j) \rightarrow$ output image pixel. Where the (i, j) indicate that the pixel is located at i -th row and j -th column. The pixel transformation of an image is done to make the model more generic and more diverse. That is if a person is with a darker complexion the eyebrow and its co-ordinates detection becomes difficult. Now the image is filled with primary colours (RGB).

Now the image is converted into binary image which first includes converting the image into greyscale, and by finding the threshold value the image pixel values are set that is if the pixel values $>$ threshold the pixel value is set to 1 else 0, hence in this way the image is converted into binary form.

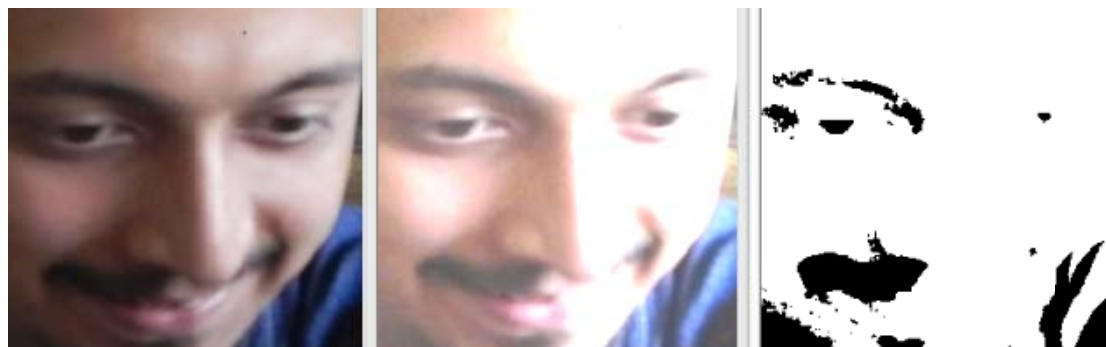


Fig 2. Image Preprocessing

2) *eyebrow Detection:*

Several approaches are applied to extract the discriminative features to learn the pattern of the different facial features. We investigate approaches based on the pixel analysis of the images which are normalized to standard scale (200 x 200) pixels. Pixel value analysis is the method of analysing an input image from the extreme left top, this involves the analysis of every pixel it encounters in every row of the normalized image.

Binary converted image is a digital image that has only two possible values for each pixel. Typically the two colours used for a binary image are black and white. These bi-tonal images has each pixels stored as a single bit of 0 or 1. The bit value 0 corresponds to black and that of 1 corresponds to white. The pixel value analysis of the binary image scans the image from the top left through each row. The position of the row and the column is recorded for the first encountered pixel with the value corresponding to 0 that is the first encountered black pixel is recorded with its coordinates, which is considered as the eyebrow tip co-ordinate of the normalized image. Figure 3 shows the normalized facial image. The image converted to binary form is used for the pixel value analysis technique, which produces the (i, j) coordinates of the eyebrow as depicted. The values of (i, j) being (20, 181) represents the 20 Th row and the 181 Th column, which is the coordinate of the tip of the eyebrow.

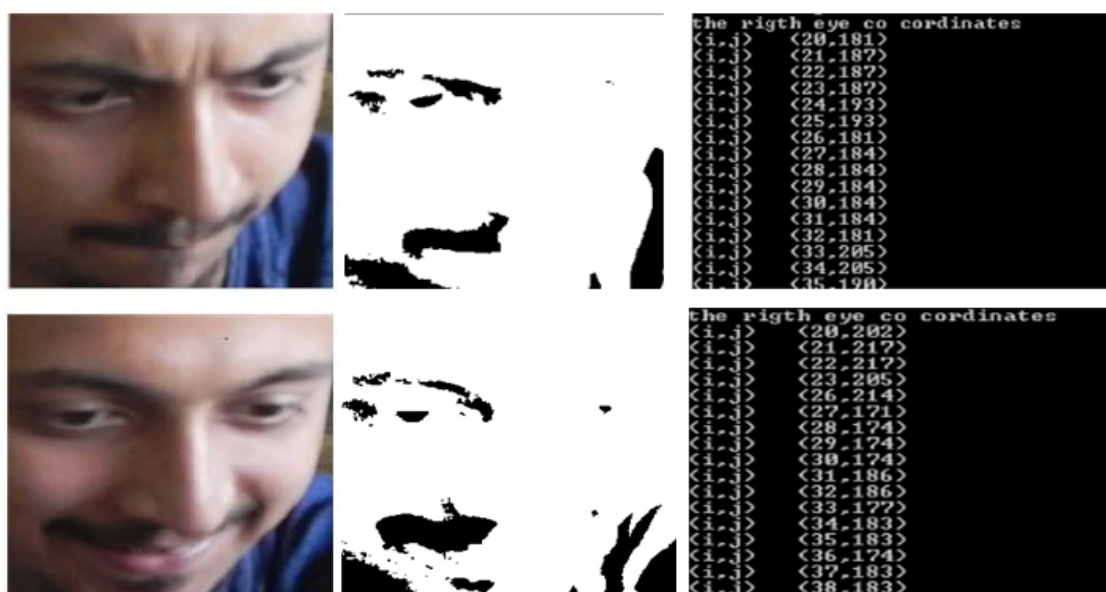


Fig. 3. Eyebrow Coordinates after preprocessing

3) *Stress Detection:*

As defined in the section Task Definition, stress detection is based on the rigid transformations shown by the eyebrows which are used as the major facial area of interest in its subsequent process of analysis. The previously

followed processes used to pre-process the input image for the eyebrow detection is further used in the stress detection methodology, which involves sub-modules of offline displacement calculation, variation of displacement and classifier which helps in stress detection.

The first sub-module offline displacement calculation, calculates the displacement of the eyebrow position using the obtained coordinates in the previous steps, with respect to its mean position. In the Fig 3, the coordinates in the first image gives the x and y co-ordinates of the eyebrow as (20, 181) and the next image gives the co-ordinates (21, 202), this sub-module gives the mean displacement of the eyebrow and this way the displacement calculation for each image with respect to first image is calculated.

The second sub-module is calculation of the variance. The eyebrow transformation in every subsequent image over a period of time is estimated by calculating its variability. And the set of images over a period of time that show variance above the threshold is classified by the sub-module as stressed over the particular time interval.

4) Deep Learning:

The deep learning module is the final module that consists of the sub-modules of training dataset, learning linear regression algorithm, model and testing the dataset for prediction. The training dataset is prepared using the obtained results in the previous modules. Using the supervised learning technique the linear regression algorithm is trained using the python package Theano, which builds a model prepared for prediction. The linear regression algorithm incorporates the gradient descent algorithm which is used to achieve optimization.

Linear regression: using the training dataset which consists of inputs of numerical deviations and the corresponding detected emotion of an individual. The linear regression model builds a linear expression which is of the form $Y = mX + c$ where in

$X = \{x_1, x_2, \dots, x_n\}$, X represents the set of numerical deviations of eyebrow movements.

$Y = \{y_1, y_2, \dots, y_n\}$, Y represents the emotion of the corresponding x_i .

With these values of x_i and y_i which forms the training dataset is used to train a model and tested using the test dataset consisting of new input values for x to predict the corresponding y values.

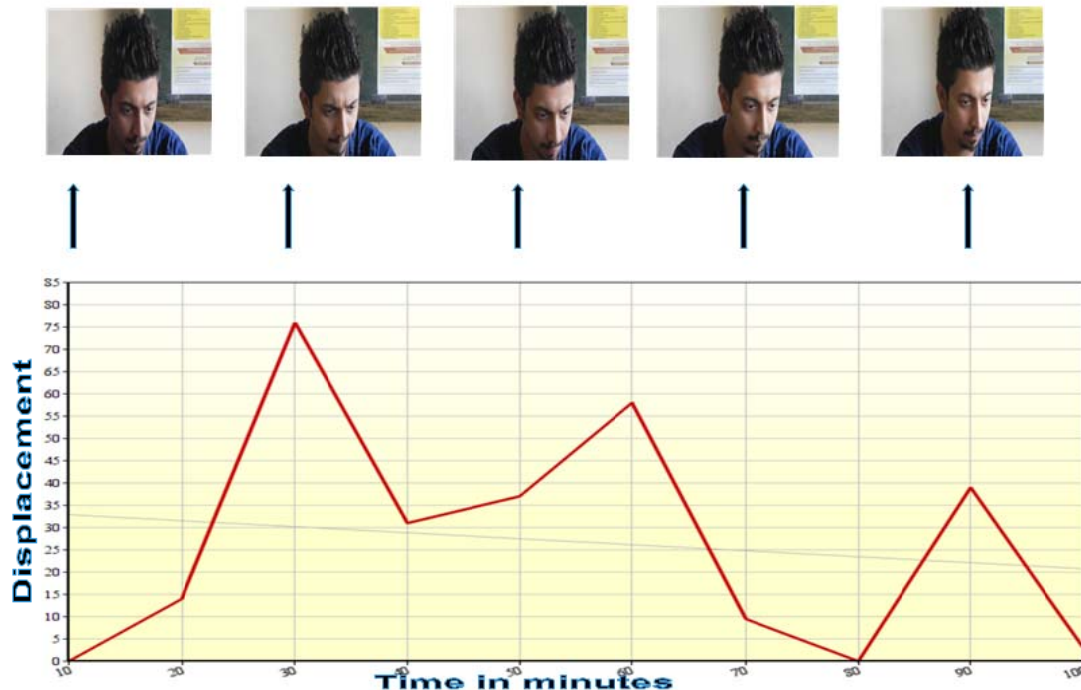


Fig. 4. Graph showing displacement of eyebrow with respect to time in minutes

IV. EXPERIMENTS AND ANALYSIS:

To evaluate the performance of the developed system, we collected data and carried out quantitative experimental evaluation. The results are discussed and justified in this section.

A. Data Collection

The dataset consists of set of image frames which are extracted from the videos recorded in the university during the lab hours. The videos were recorded during the lab hours with a camera placed on a tripod in front of the subjects. None of the recorded subjects are professional actors. The subjects during the lab hours work

continuously in front of the computer with the assigned tasks. These images form the basis of our dataset. The videos recorded during the sessions are used for the model adaption in its analysis. Video of an individual is captured during the lab hours for the duration of 3 hours. Set of 6 frames of an individual were extracted from the video every hour, having a total of 18 (6 * 3) frames during the session of three hours.

The whole dataset consists of video recording and the corresponding frames of (several individual), which is further divided into the training and the testing datasets. The training dataset consists of all the samples from the entire dataset and the remaining forms the testing dataset used to test the trained model.

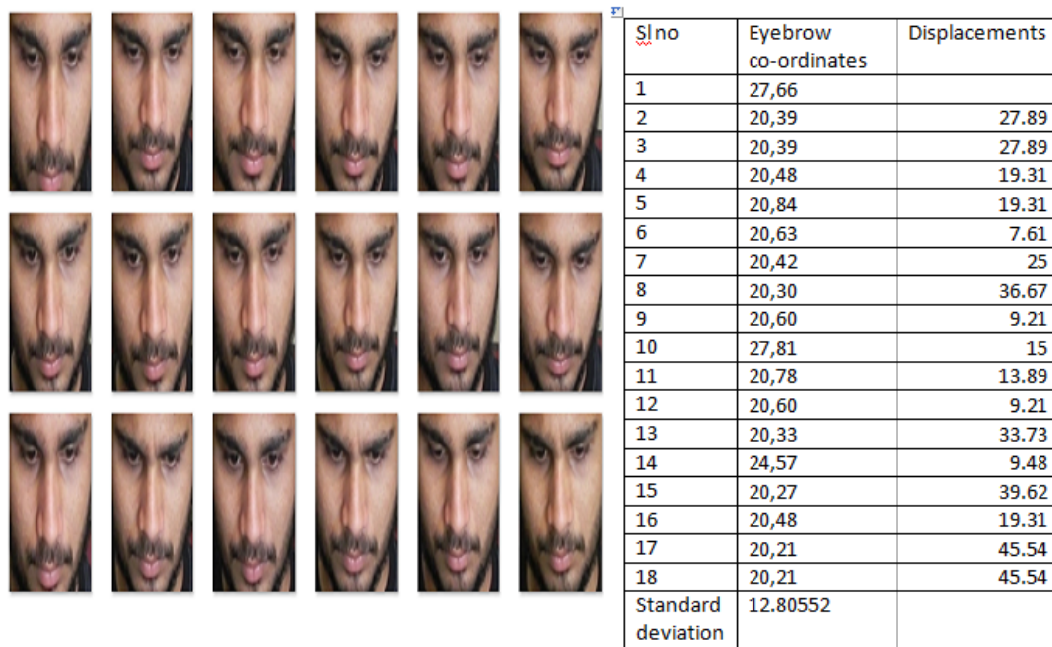


Fig. 5. Data set along with their eyebrow co-ordinates

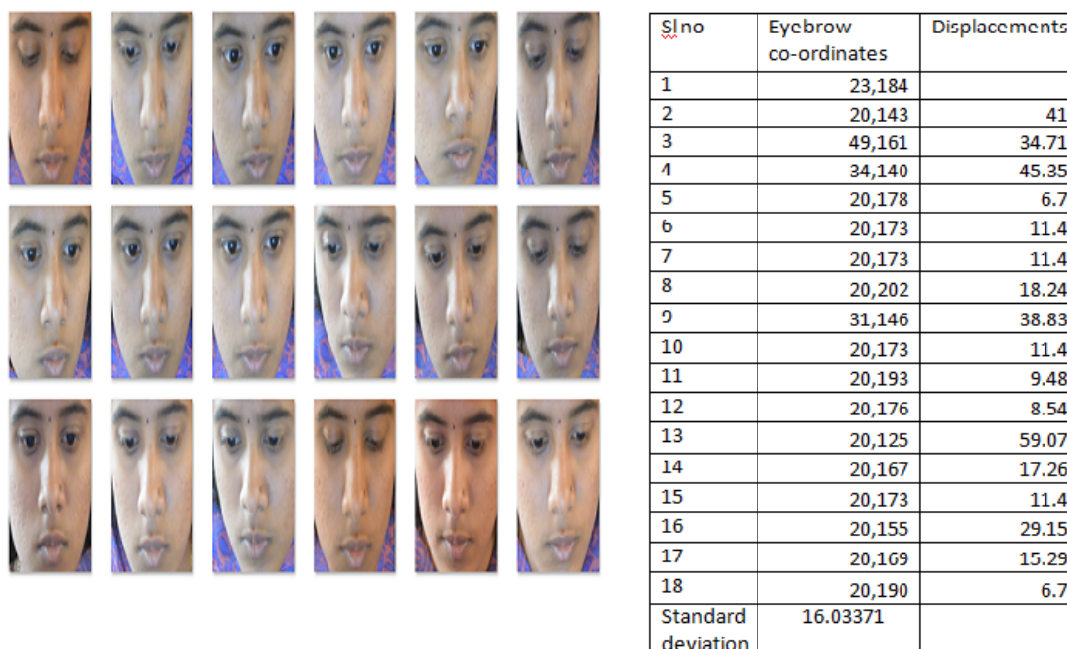


Fig. 6. Data set along with their eyebrow co-ordinates

The Fig 5 and Fig 6 shows the data set along with their eyebrow co-ordinates, displacement of eyebrow from its mean position and the standard deviation. These data sets form the database which are primarily divided into training and the testing data set. The entire dataset consists of 19 sets of images of 19 different individuals. The dataset is divides into training and the testing data sets

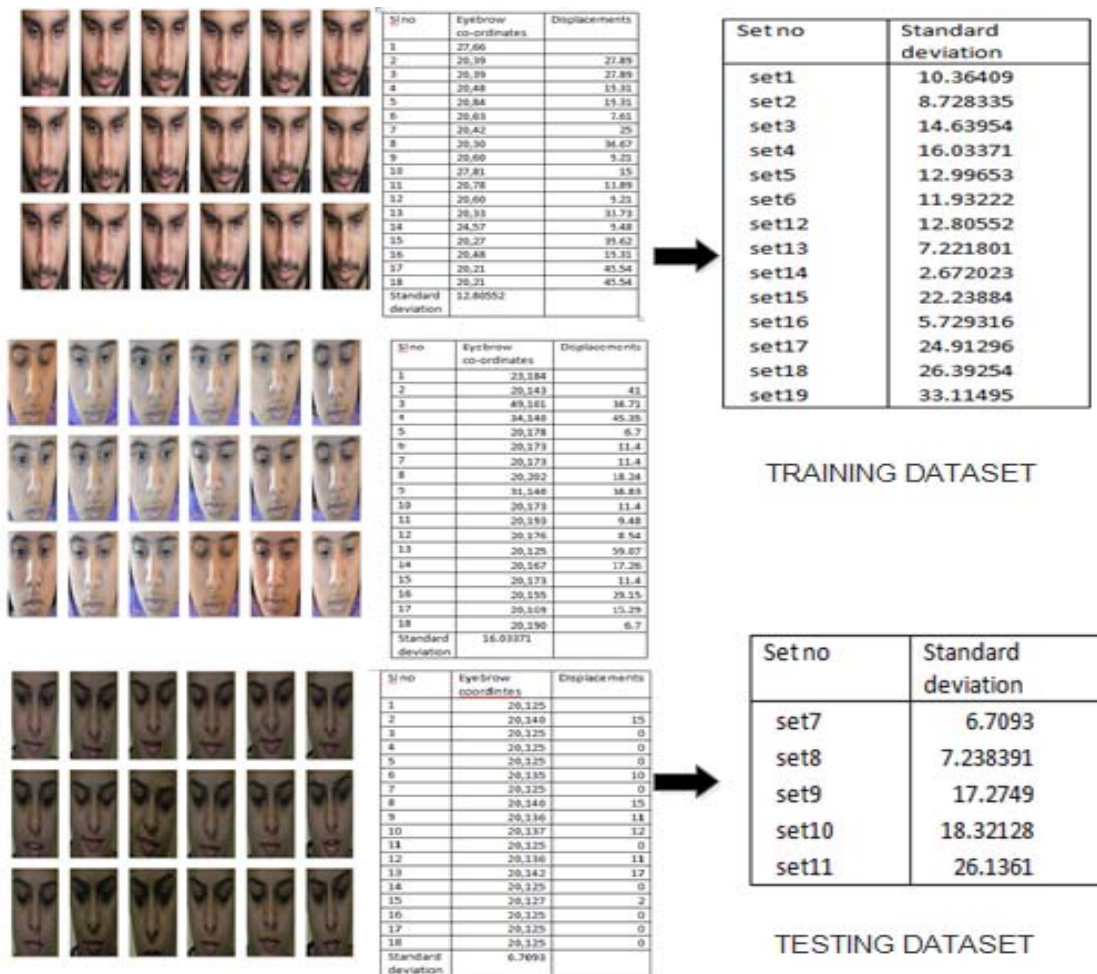


Fig 7.The Training and Testing Dataset

The training dataset is prepared using the obtained results in the previous modules. Using the supervised learning technique the linear regression algorithm is trained using the python package Theano, which builds a model prepared for prediction, is incorporated to predict the results for the stress in a person. The testing dataset being a part of the dataset is tested for the desired results using the trained model on GPU. The trained model Theano precisely predicts if a person is stressed.

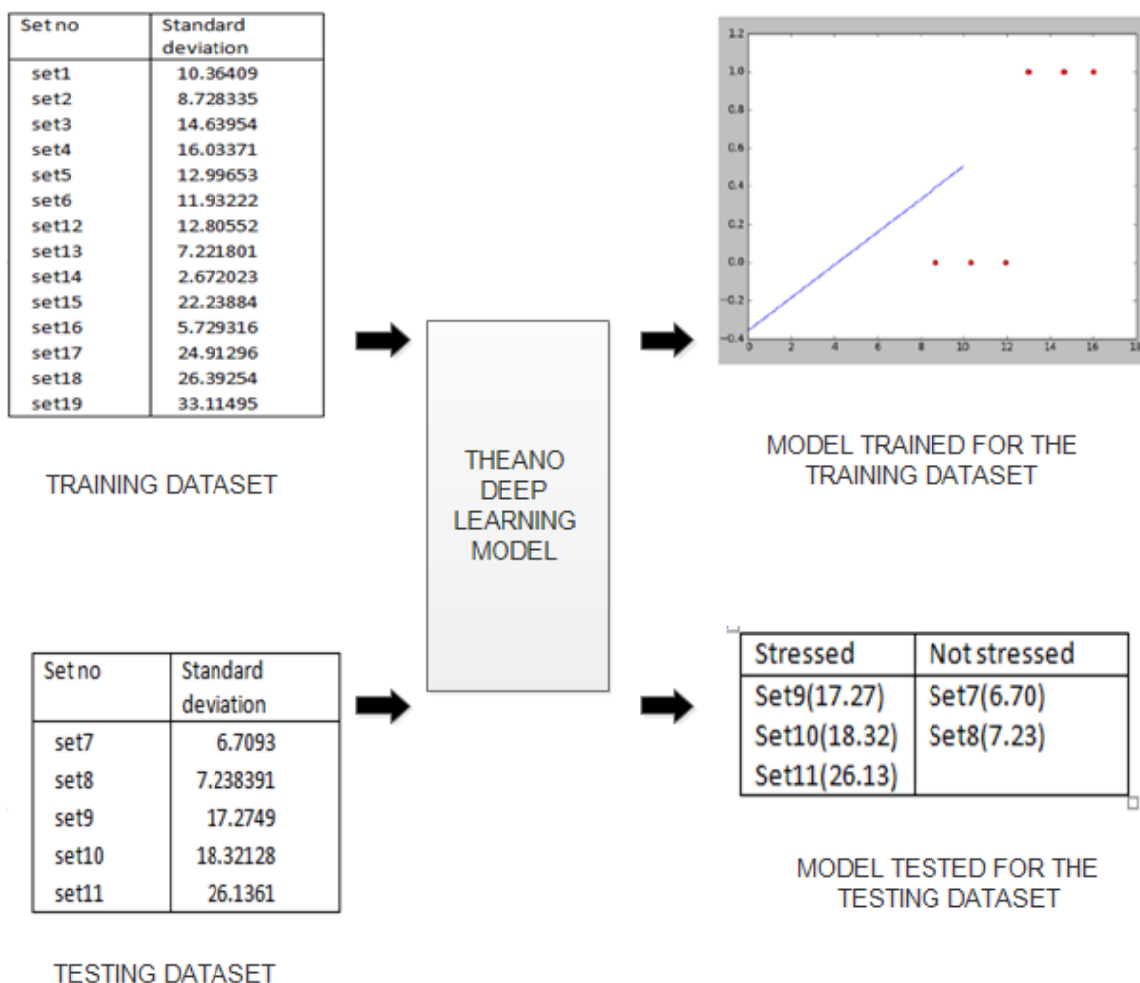


Fig 7.The Training of Deep Learning Model

The model trained using the training dataset depicts the set of points above a line as stressed and the set of points below them as not stressed. This model is tested for the set7, set8, set9, set10, set11. The deep learning model after the training predicts if the set and their corresponding person is stressed.

V. CONCLUSION AND FUTURE WORK:

We developed a monitoring system for detecting emotional stress of a person working continuously in front of computer. To assess the detection performance, we conducted experiments on 19 collected data sets each consisting of 18 images of an individual , hence total of 19x18 images analysed.

The proposed system integrates image processing and deep learning to detect stress We had collected images of each person for over a month .These images were processed in order to extract a feature (eyebrow). The obtained results from image processing with suitable inputs were used to train linear regression model and test this model with the test data set. Although the obtained results are preliminary with limited number of participants and data, the main added value of our work is achieved by enabling end users to successfully recognize his ongoing stress in order to minimize future health risks. A study on a larger population will form a part of our future work.

REFERENCES

- [1] Tomas Simon Kruez2 Iain Matthews3 Ying Yang1 Minh Hoai Nguyen2 Margara Tejera Padilla2 Feng Zhou2 Jeffrey F. Cohn1, 2 and 3 Disney Research Pittsburgh PA USA Fernando De la Torre2 1 University of Pittsburgh, 2 Carnegie Mellon University. "Detecting Depression from Facial Actions and Vocal Prosody".
- [2] ACT Australia Australian Nat. Univ., Canberra. "Eye movement analysis for depression detection".
- [3] Australia David Vandyke Human-Centred Computing Lab University of Canberra, Australian Capital Territory. "Depression Detection Emotion Classification via Data-Driven Glottal Waveforms".
- [4] Hichem Sahli Lang He, Dongmei Jiang. "Multimodal Depression Recognition with Dynamic Visual and Audio Cues".
- [5] Universite de Montreal Yoshua Bengio Dept. IRO. "Learning Deep Architectures for AI".
- [6] H2C 3J7 Canada Yoshua Bengio Dept. IRO, Universite de Montreal. Montreal (QC). "Deep Learning of Representations for Unsupervised and Transfer Learning".

AUTHOR PROFILE

Nisha Raichur Software Engineer at Tata Elxsi, ITPL main road, Hoodi Bengaluru,India

Nidhi Lonakadi Software Engineer at Accenture, Ecospace Bellandur Bengaluru,India

Priyanka Mural Software Engineer at Infosys, PocharamVillage Hyderabad India