JOURNAL OF AERONAUTICAL MATERIALS
ISSN: 1005-5053

Vol. 42, Issue-05, 2022
pp. 393-400

IMAGE PROCESSING AND MACHINE LEARNING DEPLOYED TO FACIAL EXPRESSIONS FOR THE INTENT OF EVALUATING STRESS

Dr. Gurpreet Kaur
SGGS College, Affiliated to Punjab University, Chandigarh, India

Abstract: The way in which one deals with stress is, to a significant extent, an issue of human dignity. People who have poor self-esteem are more likely to suffer from depression than their counterparts who have higher levels of this trait. For the sake of a person's comfort, it is absolutely necessary to keep an eye on the mental state of someone who is working for an extended period of time. Using advanced methods of machine learning and image processing, the purpose of our study is to identify signs of stress in professionals, particularly those working in the field of information technology. In this particular system, it consists of live detection as well as periodic analysis of staff members. As a consequence of this, it is determining the levels of both mental and physical stress in the individual in question, as well as supplying them with the appropriate treatments for stress management in the form of an ongoing survey. In the course of this investigation, a method is utilised that makes it possible to train a model and investigate the changes in how the attributes are predicted. Our method is primarily geared at the alleviation of stress, the promotion of a natural and healthy working environment for staff members, and the elicitation of optimal performance from employees while they are on the job.
Keywords: Image pre processing, Pixel transformation, Random Forest classifier, Supervised Machine Learning i.e. Deep Learning, Training dataset.

1. Introduction

In today's world, workaholism has become the norm. As a result, the demands of the workplace continue to increase. There is a lot of pressure on workers, especially in the IT sector, to create innovative new products that take use of cutting-edge technologies. In this scenario, it's not just productivity that's up, but also the employees' stress levels. Although many companies have programmes aimed at improving employees' mental health, workplace

stress is still out of hand. The onset of depression is often attributed to stress. Problems in any one of these areas might add up to a significant amount of stress for you. As a result, it contributes to a wide range of health issues, including cardiovascular disease, dementia, and more.

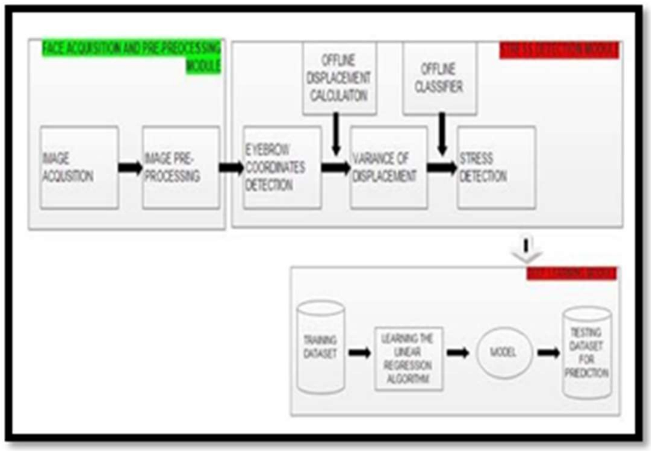

Fig.1 Module explaining the system overview

Employees may use the Stress Detection System's preventative stress management solutions, which focus on reducing stress and boosting the health of the workforce, to better deal with the causes of their stress. Books on medicine [14] conservatively estimate that stress is the root cause of between 50% and 80% of all physical disorders. Scientists have long suspected that stress is a major contributor to heart disease. Diabetes, ulcers, asthma, migraine headaches, skin problems, epilepsy, and erectile dysfunction are all conditions that might worsen under stress. These illnesses, as well as many others, are all considered psychosomatic, meaning they are either brought on by or exacerbated by psychological factors like stress. There are three main impacts of stress:

Feelings of guilt, shame, anxiety, aggressiveness, or irritation are all subjective responses to stress. People might be exhausted, tense, anxious, irritated, emotional, or isolated.

Behavioral impacts of stress are those that may be observed in a person's daily life. Accidents, drug or alcohol usage, inappropriate laughing, bizarre or argumentative conduct, extreme agitation, and binge eating are all indicators of behavioural stress.

Cognitive stress can cause a decline in cognitive functioning, faulty decision making, hasty actions, forgetfulness, and/or a heightened sensitivity to criticism.

In this study, we take a deeper dive into this issue by exploring how to use image processing and machine learning methods to understand stress patterns in the workplace.

Stress classification is an application of machine learning methods such as Random forest classifiers. The first level of detection [13] makes use of Picture Processing, with the employee's image being captured by a camera and then utilised as input. The purpose of image processing is to improve a picture or derive usable data from it by transforming the image into a digital format and then applying various operations to the digital data. Processing input from video frames as a picture, with potential outputs including both the image and features associated with it. The three main stages of processing an image are:

- This includes using picture capture tools to import the image.
- Image manipulation and analysis.
- A report or image that has been modified thanks to some sort of processing done on the input image.

2. LITERATURE REVIEW

Most programmes only aim to identify a handful of prototypical emotions, those that can be recognised by people of any race or culture. Disgust, fear, pleasure, surprise, sorrow, and anger are the six core emotions identified by [1] Since the neutral expression takes on many of the traits shared by the six other fundamental expressions of emotion, it would be best to think of it as the seventh. Variations in stance, rather than skeletal muscle activity, are mostly responsible for the wide range of neutral body postures. In normal life, we rarely experience strong feelings. Subtle changes in particular regions, such the eyebrows or eyelids, have a larger role in the expression of emotions in everyday life. For instance, one may draw their lips together in an angry expression or lower their lips in a sorrowful one. It's important for facial expression recognition algorithms to be able to pick up on even the most nuanced of changes, since this will ultimately lead to more accurate results. The methods for human coders to identify facial movement are provided by the Facial Action Coding System [2]. In order to decode Action Units, coders must first observe a series of a human subject's facial behaviour (AU). Each Action Unit represents a single gesture, such as the lifting of an eyebrow or the lowering of an eyelid, or a specific muscle contraction, such as a smile or a frown. There are 44 a.u. that make up FACS. Each action unit may alternatively be described on an intensity scale from 1 to 5. Despite the fact that Ekman and Friesen's suggested combinations of action units as descriptors of certain emotions have been integrated into FACS, the framework itself does not carry such information. The Emotional Modality FACS (EMFACS) is one such encoding scheme that is used for this purpose [3]. Expressions like grief and surprise can be coded by transforming FACS's action units into those of EMFACS or another emotion-specific system. There appears to be a difference between natural and posed (or requested) facial expressions, as recorded in many academic works [4]. The subcortical regions of the brain and the cortical motor strip are responsible for spontaneous actions, whereas the prefrontal cortex is responsible for planned activity [5]. Physiologically speaking, this makes perfect sense because of how the brain is structured. The primary characteristics that differentiate spontaneous and posed facial expressions are the actual movement that is initiated from facial muscles and the kinematics of the expression [6]. Synchronized, smooth, symmetric, constant, and reflex-like facial muscle action is a hallmark of subcortically originated (spontaneous) facial emotions. Cortically initiated (posed) facial expressions, on the other hand, are often less smooth and have more variable dynamics [7]. As such, trustworthy annotated datasets are required for the development and evaluation of such systems. Several attempts at creating such databases have been made in the literature, but it is challenging to understand all the various sources of variability in a single database. The Japanese Female Facial Expression Database (JAFFE) is one such example [8]. There are 10 distinct Japanese ladies shown here, each posing three or four times to illustrate a range of fundamental emotions, for a grand total of 213 photographs. However, unlike JAFFE, the Cohn-Kanade database [9] is commonly used for facial expression analysis, because it includes temporal information [10]. It features 100 people presenting 23 face displays with FACS annotation and basic emotion tags. It has disadvantages but is extensively used to evaluate AFER systems. The dynamics of each expression should have three states: the onset, the apex, and the offset. The Cohn-Kanade database lacks expression offset information. The Cohn-Kanade database's photos include timestamps that overlap with the subject's expression, another drawback. MMI [11] includes prepared and spontaneous face actions. It also has 600 photos and 4000 movies. FACS codes single action units, combinations, and fundamental emotions for the visuals. Additionally, profile views are given. The 3D Yin Facial Expression Database [12] was recently created. 3D models, textures, and raw model data are expressive data. It gives landmark points for face features segmentation evaluation. It has 6 primary emotions and neutral. Most AFER research groups assess approaches using databases or their own signals. This modest fragmentation on system assessment prevents comperative examination of all literature-proposed approaches.

3. Objectives

- To imagine a person's stress level by observing their facial expressions.
- To take a close look at the worker's stress levels.
- By giving answers and treatments, we hope to alleviate the tension that may arise after an individual's photographs have been analysed.

4. Methodologies

The first step in the imaging process is called "pre-processing," and it entails two different modifications of the retrieved frame. The first is a pixel transformation, while the second is a binary transformation of the updated picture. G I j) = α. The image's brightness and contrast may be adjusted with the parameters gain and bias, respectively denoted by F I j) +, > 0, and. In this case, G I j) represents the final picture pixel and F I j) represents the original image pixel. In this notation, the pixel occupies the i-th position in the j-th column of the image. It is done to make the model more generalizable and flexible, and the image is transformed pixel by pixel. Image is filled with primary colours if detecting a person's eyebrow's coordinates gets difficult due to their darker skin tone (RGB). One method of extracting pixel values in image processing is by pixel transformation. This adjustment is used to broaden the appeal of an image. The picture is transformed from its original colour to a grayscale representation. The picture threshold is determined, and pixels with values more than the threshold are converted to 1s while those with values less than the threshold are assigned 0s.
Facial Expression Recognition:

The six basic steps of a typical camera-based FER(Facial Expression Recognition) system are: pre-processing, feature extraction, classification, measure similarity, and retrieval of data with the minimum distance or displacement. Beginning with an input image, the facial region is identified and pre-processed so that recognisable features, such the eyes and nose, may be extracted. Later, characteristics are retrieved from the face to aid in the recognition phase, and in the third step, different facial emotions are categorised. The fourth stage involves quantitative analysis of similarities. Lastly, eyebrow displacement may be estimated using the standard deviation as a proxy for stress detection. The eyebrows are a particularly informative part of facial images because they may convey a wide range of emotions, including disdain, fear, and fury.

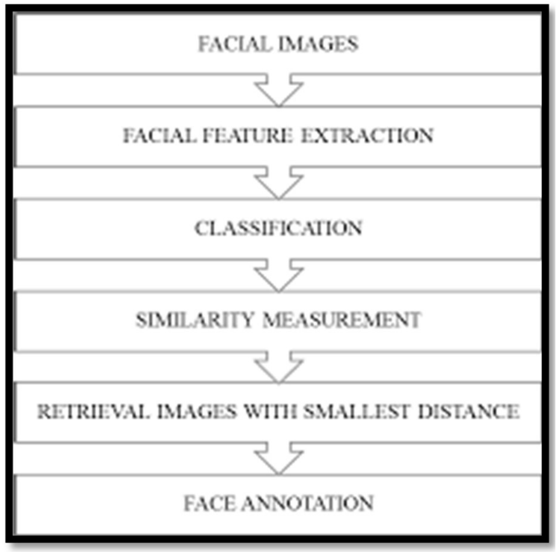

Fig.2 Steps to analyze facial expression

In Fig. 2, we can see the numerous processes that are employed in order to draw conclusions from a comparison of a face picture with different types of face annotation.

The Movement of the Eyebrow There are a variety of methods for isolating distinguishing characteristics in order to study the underlying structure of one's own face. Objective: explore techniques based on pixel analysis of photos scaled to a standard resolution of 200 by 200. Analyzing an input image using pixel value analysis entails examining every pixel in every row of the normalised image, starting at the top left corner.

Images that have been transformed to binary only allow for two possible values per pixel. The two standard tones of a binary picture are black and white. Each pixel in these binary graphics is represented by a binary value of 0 or 1. Black is represented by the bit value 0 and white by 1. Analysis of the binary image's pixel values proceeds from top left to bottom right throughout each row. When calculating the eyebrow tip co-ordinate of the normalised picture, It is necessary to keep a record of the position of the row and column for the first pixel that is found to have a value of 0 associated with it. When a black pixel is discovered, this state of affairs always results. By utilising the approach of pixel value analysis in conjunction with the binary image, we are able to extract the I j) coordinates of the eyebrow in the format given.

Fig.3 Eyebrow Coordinates

Stress may be identified in fig. 3 by measuring eyebrow displacement and then categorising the results into high, moderate, or low categories.

Random Forest classifier: An RF classifier is a type of ensemble learning strategy that is built from a collection of decision trees, each of which is trained using bootstrap aggregation or bagging and is therefore a random forest. A random forest (RF) is a classifier that is resistant to over fitting; it outperforms SVM and AdaBoost-based algorithms because it uses randomization to choose subsets and features for the trees to develop from. An RF decision tree uses bagging to choose subset S of training samples for use in the training job. In a top-down induction, a binary decision tree is first created from the leaf node and then branches out from there. The split function f(v) consists of a collection of randomly picked feature vectors v and a threshold. Its purpose is to divide the subset Qi that is located at the i-th node into the subsets QLi and QRi. Iteratively generating the feature vectors and threshold value allows one to arrive at the optimal split function more quickly. There are many other conceivable combinations of split function and threshold, but the system chooses the one that generates the most amount of usable data for the target node. This might be any one of a number of different combinations. This process of separating nodes takes place several times until either the maximum depth is reached or there is no longer any new information acquired. After all of the above steps have been completed, a leaf node will have a posterior probability and class distribution p(c|l) for each class.

While RF outperformed KNN, the previous method used to categorise the data, when the training sample size was small, the two methods achieved similar results when the sample size was medium.

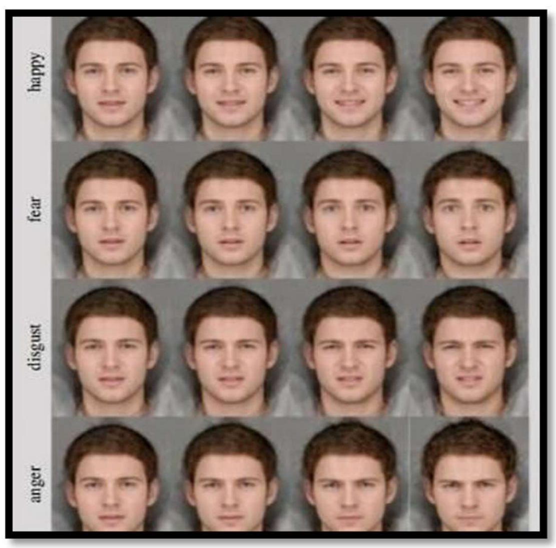

Fig.4 training dataset

Finally, we have deep learning, which is split down into the sub-modules of training datasets, learning linear regression technique, model, and acquiring outcomes of prior models. The R programming language is used to train a linear regression algorithm utilising the supervised learning approach; this algorithm will evaluate data and create a prediction model. The optimization technique known as gradient descent is built into the linear regression method.

Linear regression analyses the association between numerical outliers and the associated identifiable emotions of a person using the outliers and emotions as inputs. An expression of the form Y=mX+c can be derived from a linear model.
where X = x1, x2,..., xn is a collection of numerical variations of eyebrow motions.

If you write Y= y1,y2,...,yn, Y will indicate the feeling associated with each xi. The xi and yi pairs in the training dataset are used to teach a model to predict fresh input values for x, and the model's performance is evaluated using the test dataset. Moreover, additional features, such as the lips, nose, and cheeks, can be analysed to determine a person's stress level.

5. Conclusion

Employee stress may be predicted using a system that monitors the photos of logged-in users, providing a high level of security. Standardized techniques for converting and processing pictures are applied to the acquired images in order to identify the user's stress based on a predetermined time period. Then, the system will conduct an analysis of the stress levels by making use of Machine Learning algorithms such as random forest, which provide accuracy in results of different datasets, and the R programming language, which generates results that are more efficient than ever before and are based on things like mood, temperature, and so on. Future studies that can't be conducted using the current methods might benefit from a different perspective in the form of side-on photographs. In addition, filters can be applied to the photographs to improve the quality and reliability of the results.

References

1. Ankita Patil1, Rucha Mangalekar2, Nikita Kupawdekar3, Viraj Chavan4, Sanket Patil5, Ajinkya Yadav6 “stress Detection in IT Prfessionals by Image Processing and Machine learning” in IJRESM vol.3 Issue 1 pp-121-123,January 2020.
2. G. Giannakakis, D. Manousos, F. Chiarugi, “Stress and anxiety detection using facial cues from videos,” Biomedical Signal processing and Control”, vol. 31, pp. 89-101, January 2017.
3. ‘Nisha Raichur, Nidhi Lonakadi, Priyanka Mural, “Detection of Stress Using Image Processing and Machine Learning Techniques”, vol.9, no. 3S, July 2017.
4. T. Jick and R. Payne, “Stress at work,” Journal of Management Education, vol. 5, no. 3, pp. 50-56, 1980.
5. 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".
6. U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2018, pp. 1-4.
7. Bhattacharyya, R., & Basu, S. (2018). Retrieved from ‘The Economic Times’.
8. OSMI Mental Health in Tech Survey Dataset, 2017
9. https://www.kaggle.com/qiriro/stress

10. ACT Australia Australian Nat. Univ., Canberra. "Eye movement analysis for depression detection".
11. Australia David Vandyke Human-Centred Computing Lab University of Canberra, Australian Capital Territory. "Depression Detection Emotion Classification via Data- Driven Glottal Waveforms ".
12. Hichem Sahli Lang He, Dongmei Jiang. "Multimodal Depression Recognition with Dynamic Visual and Audio Cues ".
13. University de Montreal Yoshua Bengio Dept. IRO. "Learning Deep Architectures for AI".
14. H2C 3J7 Canada Yoshua Bengio Dept. IRO, University de Montreal. Montreal (QC). "Deep Learning of Representations for Unsupervised and Transfer Learning”