Facial Expression Recognition System for Stress Detection

Meenaksi Chowdary Para
School of Computer Science
Texas A&M University
Corpus Christi, TX
mpara@islander.tamucc.edu

Shivani Mothe
School of Computer Science
Texas A&M University
Corpus Christi, TX
smothe@islander.tamucc.edu

Manideep Reddy Pati
School of Computer Science
Texas A&M University
Corpus Christi, TX
mpati@islander.tamucc.edu

I. Introduction

Demanding jobs are a significant cause of stress in people. Situations like frequent exposure to danger, short deadlines, rigorous tasks or even repetitive tasks are some stress originators.

Nearly one in three workers in Europe and the United States report that they are affected by stress at work. Work-related stress, depression, and anxiety can result in reduced work performance and absenteeism, costing an estimated 3% to 4% of gross national product (Dewa & Hoch, 2015). Also, about 61% of European institutions participating in the 2019 EU-OSHA study (EU-OSHA, 2019) reported that a reluctance to talk openly about these issues seems to be the main difficulty for addressing psychosocial risks [1].

There is evidence that stress conditions are both preventable and treatable in the workplace and that workers who receive treatment are more likely to be more productive (Carolan et al., 2017). Hence, nonintrusive stress sensing tools that continuously monitor stress levels, with a minimal impact on a workers' daily lives, could be used to automatically initiate stress-reduction interventions. In stressful work settings, these applications could not only lead to more timely and reduced-cost interventions, but also to more productive environments where workers could better manage their workload.

Traditional stress detection relies on psychological questionnaires or professional psychological consultation. As the results of questionnaires depend largely on the answers given by individuals, the stress measure is quite subjective. When people choose to express their psychological states with reservations, the result scale would be biased. To overcome the limitations of the questionnaire surveys, the methods of automatically detecting stress by sensing an individual's physical activities through wearable devices such as mobile phones with embedded sensors or based on physiological signals such as heart rate variability HRV, electrocardiogram ECG, galvanic skin response GSR, blood pressure, electromyogram, electroencephalogram EEG, etc. from dedicated sensors have been developed. While these methods are able to objectively sense people's stress states, they usually demand wearable equipments and sensors, which could hardly realize contact-free measurement [2] [3].

Currently, the ubiquitous deployment of contact-free video cameras in surroundings, together with the rapid progress of data collection and analysis techniques, offers us another channel to detect one's stress based on image sequences captured from a monitoring video camera. Compared with previous sensory devices, the later offers the following three benefits. First, it is more convenient, particularly in places like schools, hospitals, and restricted areas like prisons, where no carry-on devices are needed or allowed. Second, it has a very long standby time and can easily reach mass audience at a very low-cost. Third, the continuous frames it captures enable us to grasp and analyze people's stressful states more naturally without interference of artificial traits and factors.

The aim of this study is to leverage contact-free video cameras for stress detection.

II. Related Work

For many years, researchers have been studying ways to automatically detect stress. Various methods have been explored, ranging from intrusive approaches like saliva or blood tests to less intrusive approaches that involve image collection.

One such study (Gao et al., 2014) involved mounting a camera inside a car dashboard to collect images of the driver's face for stress detection. These images were then classified using Support Vector Machines (SVMs) trained on public facial expression datasets, with an algorithm counting the number of anger and disgust classifications within a time window. If the number exceeded a specific threshold, the driver would be considered stressed. The best classifier was trained using not only public dataset images, but also images of subjects posing for stress classification. This allowed the models to adapt to the subjects' unique facial expressions, resulting in an accuracy of 90.5% for stress classification [4] [5] [6].

In another study (Maaoui et al., 2015), a system was developed that used a computer's webcam to collect Remote Photoplethysmography (rPPG) signals for stress detection. These signals were translated into a sinusoidal wave representing the heart rate, and an SVM classifier achieved the best results with 94.40% accuracy.

The group (Giannakakis et al., 2016) developed a system that detected stress/anxiety emotional states through video-recorded facial cues. They used techniques such as Active

Appearance Models, Optical Flow, and rPPG to extract relevant features for stress classification. The best classification accuracy of 87.72% was achieved with a K-NN classifier.

Lastly, in the study (Viegas et al., 2018), a system was proposed that detected signs of stress through Facial Action Units (FAUs) extracted from videos. They performed binary classification using several simple classifiers on FAUs extracted in each video frame and were able to achieve up to 74% accuracy in subject independent classification and 91% accuracy in subject dependent classification [7] [8].

III. Proposed Method

The system described aims to detect and notify users when they show signs of stress by analyzing their facial expressions through video images. The program is designed to run in the background, continuously monitoring the user's facial expressions and notifying them when they display signs of stress. One of the significant advantages of this program is that it can easily access the user's webcam when they are working with computers [9] [10] [11].

The system comprises several modules that work together to achieve its objective. The first module captures real-time images from the computer's webcam and sends them to the second module. The second module identifies the user's face's location and crops it using a Haar-like feature selection technique. The face is then resized and normalized to 299x299 pixels, allowing for more accurate facial expression analysis. The third module, the Emotion Classification module, uses a trained classification model to classify the face and return a list of seven probability scores, one for each facial expression.

The fourth module is responsible for stress assessment. It receives the highest probability facial expression classification from the Emotion Classification module and records various classifications made over time. Based on the parameters set by the user, the system will determine whether the user is under stress. The program only requires three parameters to function correctly: the frequency of the program, the time window, and a threshold. The frequency determines the interval between each image extraction, while the time window specifies the period of past classifications that will be considered in the stress assessment process. The threshold indicates the percentage of negative emotions needed to determine if the user is under stress within the time window.

For instance, if the time window is set to 15 minutes, and the threshold is 75%, the system determines that the user is under stress if 75% or more of the classifications in the past 15 minutes are for stressful emotions. Upon detecting stress, the system displays a notification alerting the user of the fact. If the user wishes, they can confirm this notification. Additionally, the program offers the option of saving the images collected by the webcam on disk. The images that the model positively classifies as stressful are labeled with stress and its timestamp. These images serve two purposes: first, to create a dataset of images labeled with stress/non-stress to aid future research, and second, to make these images available for expert analysis.

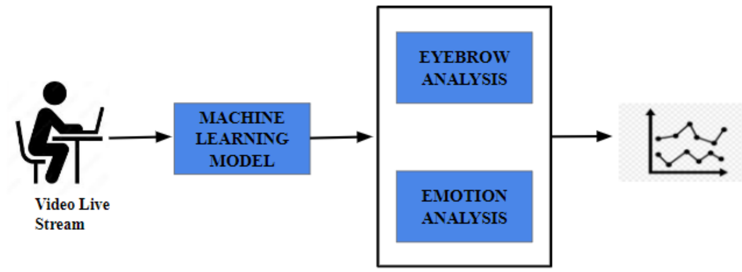

Fig. 1. Architecture of the proposed model

IV. Experiments

A. Datasets

In training on FER2013, we adhere to the official training, validation, and test sets as introduced by the ICML. FER2013 consists of 35888 images of 7 different emotions: anger, neutral, disgust, fear, happiness, sadness, and surprise. A Kaggle forum discussion held by the competition organizers places human accuracy on this dataset in the range of 65 – 68%. To account for the variability in facial expression recognition, we apply a significant amount of data augmentation in training. This augmentation includes rescaling the images up to ± 20% of their original scale, horizontally and vertically shifting the image by up to ± 20% of its size, and rotating it up to ± 10 degrees. Each of the techniques is applied randomly and with a probability of 50%. After this, the image is then cropped to a size of 40×40, and random portions of each of the crops are erased with a probability of 50%. Each crop is then normalized by dividing each pixel by 255.

Expression	Number of Images
Angry	491
Disgust	55
Happy	879
Sad	594
Fear	528
Surprise	416
Neutral	626

Fig. 2. FER2013 Dataset

B. Baselines

We explored transfer learning, using the Keras VGG-Face library and each of ResNet50 and VGG16 as our pre-trained models. To match the input requirements of these new networks which expected RGB images of no smaller than

197x197, we resized and recolored the 48x48 grayscale images in FER2013 during training time.

ResNet50 is the first pre-trained model we explored. ResNet50 is a deep residual network with 50 layers. It is defined in Keras with 175 layers. After 5 epochs of training, we achieved an accuracy of 85.70% on the test set.

Although much shallower than ResNet50 and SeNet50 with only 16 layers, VGG16 is more complex and has many more parameters. We kept all pre-trained layers frozen and added two FC layers of size 4096 and 1024 respectively with 50% dropout. After 5 epochs of training with the Adam optimizer, we achieved an accuracy of 85.73% on the test set.

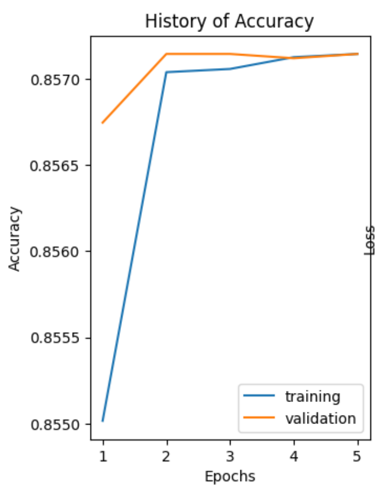

Fig. 3. Model Accuracy of ResNet50

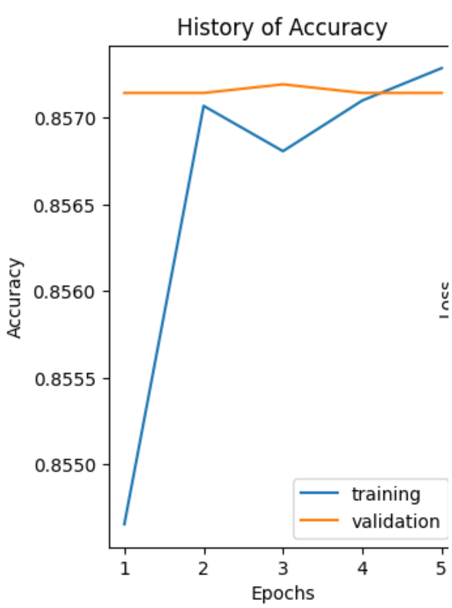

Fig. 4. Model Accuracy of VGG16

	ResNet	VGG16
Accuracy	85.70	85.73
Loss	1.80	1.825
AUC	0.65	0.65
Precision	0	0
F1 Score	0	0

Fig. 5. Comparision metrics for 2 models i.e., VGG16, ResNet50

C. Results

:

After successfully saving the VGG16 model, we got an anomaly graph for a live video stream of a person showing different emotions. The scores of stress are plotted on the graph as shown in Fig 6.

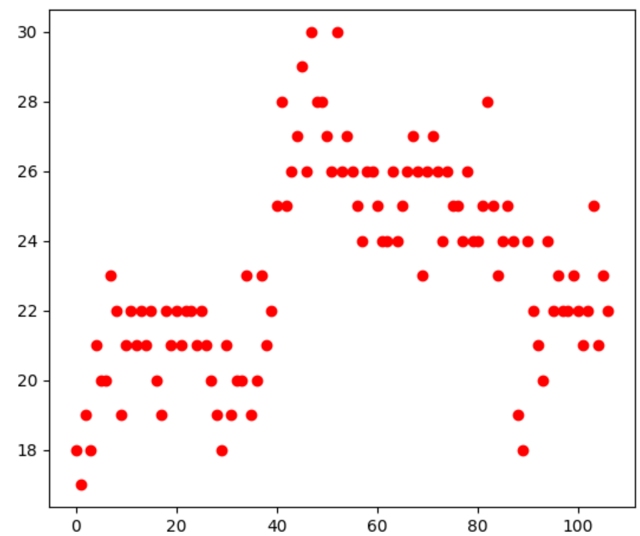

Fig. 6. Anomaly graph of a person stress levels

V. Conclusions

We have created a system that can capture real-time images of a user's face and use a facial expression classifier to determine if the user is showing signs of stress. If the user is stressed, the system will alert them. To develop the classification model, we used transfer learning and fine-tuning techniques with pre-trained networks such as VGG16, VGG19, and Inception-ResNet V2. However, the lack of a dataset specifically classified in terms of stress or non-stress meant that the accuracy of the best model was 92.1% based on the association between facial expressions and stress. To improve the accuracy of the model and support the relationship between facial expressions and stress, we plan to collect feedback from users on the alerts our system sends them.

In addition to collecting feedback, we also plan to improve the classification models by training them with more data collected from our users. We will also consider migrating the classification module to a server that can take advantage of centralized processing with graphics cards, reducing the impact on users' devices. These proposals for future work will increase the system's confidence in correctly classifying stressful situations and help establish a stronger association between negative emotions and stress.

REFERENCES

[1] Stephany Carolan, Peter R Harris, and Kate Cavanagh. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. *Journal of medical Internet research*, 19(7):e271, 2017.

[2] Carolyn S Dewa and Jeffrey S Hoch. Barriers to mental health service use among workers with depression and work productivity. *Journal of occupational and environmental medicine*, 57(7):726, 2015.

[3] David F Dinges, Robert L Rider, Jillian Dorrian, Eleanor L McGlinchey, Naomi L Rogers, Ziga Cizman, Siome K Goldenstein, Christian Vogler, Sundara Venkataraman, and Dimitris N Metaxas. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, 76(6):B172–B182, 2005.

[4] Paul Ekman. What scientists who study emotion agree about. *Perspectives on psychological science*, 11(1):31–34, 2016.

[5] Paul Ekman and Wallace V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*, volume 10. Ishk, 2003.

[6] Hua Gao, Anil Yüce, and Jean-Philippe Thiran. Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 5961–5965. IEEE, 2014.

[7] Giorgos Giannakakis, Matthew Pediaditis, Dimitris Manousos, Eleni Kazantzaki, Franco Chiarugi, Panagiotis G Simos, Kostas Marias, and Manolis Tsiknakis. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31:89–101, 2017.

[8] Jason C Hung, Kuan-Cheng Lin, and Nian-Xiang Lai. Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing*, 84:105724, 2019.

[9] Choubeila Maaoui, Frederic Bousefsaf, and Alain Pruski. Automatic human stress detection based on webcam photoplethysmographic signals. *Journal of Mechanics in Medicine and Biology*, 16(04):1650039, 2016.

[10] Lucio Ciabattoni, Francesco Ferracuti, Sauro Longhi, Lucia Pepa, Luca Romeo, and Federica Verdini. Real-time mental stress detection based on smartwatch. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pages 110–111. IEEE, 2017.

[11] CK Yogesh, Muthusamy Hariharan, R Yuvaraj, Ruzelita Ngadiran, Sazali Yaacob, Kemal Polat, et al. Bispectral features and mean shift clustering for stress and emotion recognition from natural speech. *Computers & Electrical Engineering*, 62:676–691, 2017.