Towards Independent Stress Detection: a Dependent Model using Facial Action Units

Carla Viegas
Language Technologies Institute
Carnegie Mellon University
Universidade NOVA Lisboa
Pittsburgh, USA
cviegas@cs.cmu.edu

Shing-Hon Lau
Machine Learning Department
Carnegie Mellon University
Pittsburgh, USA
shinghol@cs.cmu.edu

Roy Maxion
Machine Learning Department
Carnegie Mellon University
Pittsburgh, USA
maxion@cs.cmu.edu

Alexander Hauptmann
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, USA
alex@cs.cmu.edu

***Abstract*—Our society is increasingly more susceptible to chronic stress. Reasons are daily worries, workload, and the wish to fulfil a myriad of expectations. Unfortunately, long-exposure to stress leads to physical and mental health problems. To avoid the described consequences, mobile applications have been studied to track stress in combination with wearables. However, wearables need to be worn all day long and can be costly. Given that most laptops have inbuilt cameras, using video data for personal tracking of stress levels could be a more affordable alternative. In previous work, videos have been used to detect cognitive stress during driving by measuring the presence of anger or fear through a limited number of facial expressions. In contrast, we propose the use of 17 facial action units (AUs) not solely restricted to those emotions. We used five one-hour long videos from the dataset collected by Lau [1]. The videos show subjects while typing, resting, and exposed to a stressor, being a multitasking exercise combined with social evaluation. We performed binary classification using several simple classifiers on AUs extracted in each video frame and were able to achieve an accuracy of up to 74% in subject independent classification and 91% in subject dependent classification. These preliminary results indicate that the AUs most relevant for stress detection are not consistently the same for all 5 subjects. Also in previous work, using facial cues, a strong person-specific component was found during classification.**

***Index Terms*—affect recognition, stress detection, facial action units, person independent model, person dependent model**

I. Introduction

Chronic stress is affecting an increasing number of working adults. Long exposures to stress not only affect the physical health but also have a negative impact to the emotional state and psychological state of mind. Consequences of long-term stress on workforce are already visible in companies' revenues. Accidents, absenteeism, and diminished productivity amongst others, are causing costs of $182 Billion [2]. Ebert et al. [3] showed that it is more cost-efficient to intervene before chronic stress causes health and mental problems. Detecting stress levels and their exposure time has gained interest not only from the medical point of view but also from companies which want to improve their working environment [4], [5].

Different techniques exist to detect stress by measuring physical reactions of the body. Through saliva or blood tests it is possible to measure stress hormones. Also the heart's change of behavior during stress can be detected by measuring the heart rate, heart rate variability, and blood volume pulse. Also skin temperature, pupil dilation, and respiration can reveal the presence of stress. In recent years, voice features, facial expression, eye gaze, and blink rates have gained interest as their use as indicators of stress only require video recording. An overview of the usability of the mentioned techniques at workplaces has been evaluated in [6] .

Besides detecting stress at the workplace, mobile applications have been developed to evaluate the stress all day long. In [7] physical monitoring was combined with mental stress monitoring using a wristband and sensors of the smartphone in use. Clarke et al. [8] developed a smartphone application that besides detecting stress using a Microsoft Band 2, also suggests micro interventions. However, there are several challenges that remain. Most of the methods counted require invasive gathering of the data or wearable devices, such as, wrist bands or chest bands which can be expensive. Another challenge is that people deal with stressful situations differently. Methods that are based only on context are not stable as the reaction towards potentially stressful situations is different. Additionally, stress is not always negative. Events such as weddings, or graduations cause positive stress (eustress) and not distress. Measuring only physiological signals might not be sufficient in order to make that distinction.

In this work we present a method able to detect mental

This work was supported by a fellowship from the Portuguese Foundation for Science and Technology through the CMU Portugal Program and the BioVisualSpeech project (grant CMUP-ERI/TIC/0033/2014).

978-1-5386-7021-7/18/$31.00 © 2018 IEEE

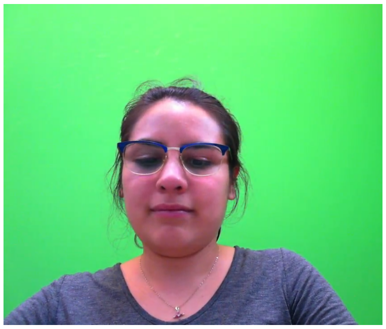


Fig. 1: Extracted frame from one video during typing phase.

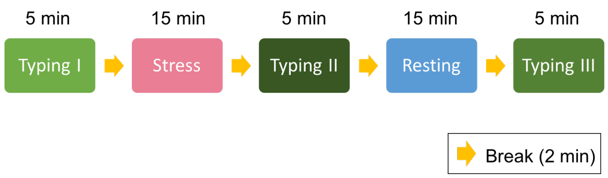


Fig. 2: Phases of the experiment recorded on video.

stress using facial expressions detected only on video. The advantage of using video for personal stress detection is the easy accessibility to webcams when working on computers. In contrast to previous work [9], [10], where stress was detected using facial expressions that are known to occur with fear, here no restrictions are used. We extracted 17 different Action Units (AUs) from upper-level to lower-level face frame-wise. We were able to distinguish stress from non-stress situations with an accuracy of 74% for person independent classification using leave-one-subject-out (LOSO). For person dependent classification we reached an accuracy of 91% using 5-fold cross-validation. Besides detecting stress through video during work for personal tracking, this method can also be useful for security applications, such as, ATM surveillance. It can also be used as a feedback tool to monitor stress during interview training as well as public speaking.

In the following section, we provide a review of existing techniques for stress detection. In Section III the dataset as well as the recording setup and the experimental protocol will be introduced. Section IV describes different experiments, with their results presented in Section V. In Section VI and VII we share our conclusions and plans for future work.

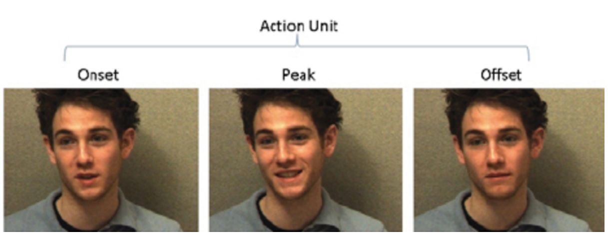


Fig. 3: Left to right, evolution of an AU 12 (involved in smiling), from onset, peak, to offset. Figure from [11].

II. Related Work

Stress is a natural process which was developed during evolution to maintain us humans alive in dangerous situations. Stress sharpens our senses and provides extra energy to react. In alerting situations we experience acute stress, but only for a short period of time. Similar in duration is eustress, which is stress experienced in positive situations, such as, graduations or weddings. Chronic stress, however, is often unnoticed as it builds up due to daily concerns, e.g., paying bills and family obligations. Untreated chronic stress can lead to serious heart health conditions, anxiety, depression, weakened immune system and obesity [12].

Chronic stress is not only a concern from a medical perspective, but it is becoming an increasing cost generator for companies. Work-related stress is estimated to cause costs of $187 Billion Dollars only in the USA [2]. In order to help employers improve their working environment, companies such as Linkura [5] offer automatic stress detection of employees and strategy plans to improve the situation.

A variety of signals have been studied for stress detection. In [13], for instance, electrocardiogram, electromyogram, galvanic skin response (GSR) and respiration were measured to detect stress during driving. As GSR provides direct insights into the autonomous emotion regulation, it is one of the most used physiological signals in stress detection [14]–[16]. Also speech has been used to detect stress by using different features such as MFCC and pitch [17], [18]. Hernandez et al. analyzed behavioral changes using a pressure-sensitive keyboard and a capacitive mouse [19]. His results show that in stressful conditions, participants increase their typing pressure and the contact surface with the mouse. Although smartphones in combination with wearable devices are ideal to track stress continuously, average accuracies vary from 71% to 75% for binary classification of high and low stress conditions for user-specific models [20]–[22].

Given stress is related to emotions, also facial cues have been used to detect stress. Lerner et al. used facial expressions that are linked to fear and anger as features [9]. She found that people react differently to the same stressor (stimulus causing stress), some showing fear and others anger. Also in [10] facial expressions were used to detect whether a person was stressed during driving or not by comparing the presence of anger and disgust, against the remaining emotions. In a similar study, cognitive load during driving was detected using facial AUs and the correlation between them [23].

In this work we want to evaluate if a combination of different Action Units (AUs) from the Facial Action Coding System (FACS) [24] can be used to indicate the presence of mental stress. To the best of our knowledge, this work is the first to use a variety of facial action units to detect mental stress, without restricting them to those active during fear or anger. We used 17 different AUs to detect mental stress caused by a multitasking exercise combined with social evaluation. We used different simple classifiers to perform a subject dependent and independent analysis.

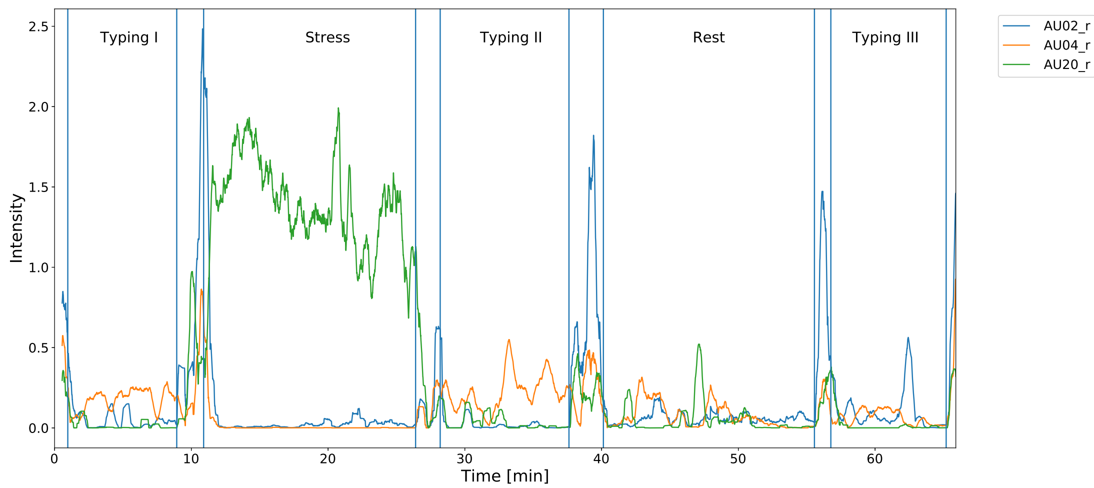
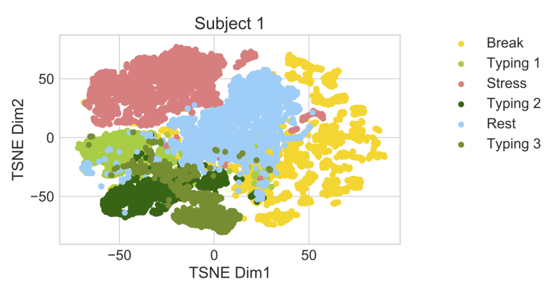
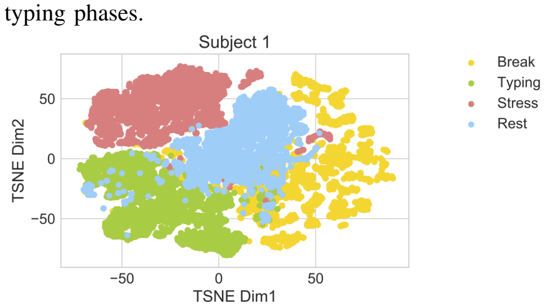


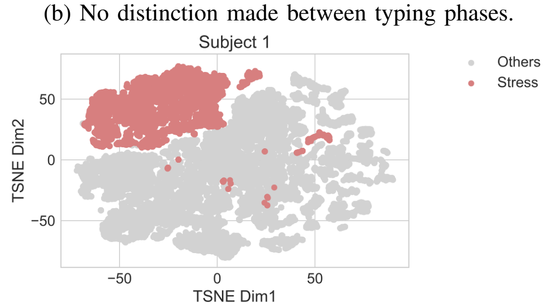
Fig. 4: Plot showing AU2, AU4, AU20 of one subject during the entire video. Start and ending of each phase are marked by vertical lines. In between the main phases the subjects were given a short break. However, the behaviour of the shown AUs was different for different subjects. Nevertheless, correlations of different AUs with the different phases were visible.



(a) All classes shown by distinguishing the different typing phases.



(b) No distinction made between typing phases.



(c) Binary problem visualization, showing stress and remaining phases grouped together.

Fig. 5: Visualization of the data of one subject using tSNE.

TABLE I: Action Units (AUs) used as features. They provide intensity information (I) which varies between 0-5. or presence/absence information (P) with a binary value.

AU	Description	Prediction
AU1	Inner brow raiser	I
AU2	Outer brow raiser	I
AU4	Brow lowerer	I
AU5	Upper lid raiser	I
AU6	Cheek raiser	I
AU7	Lid tightener	P
AU9	Nose wrinkler	I
AU10	Upper lip raiser	I
AU12	Lip corner puller	I
AU14	Dimpler	I
AU15	Lip corner depressor	I
AU17	Chin raiser	I
AU20	Lip stretcher	I
AU23	Lip tightener	P
AU25	Lips part	I
AU26	Jaw drop	I
AU45	Blink	P

III. Data

For this work, videos from five subjects collected by Lau [1] were used. The videos were collected with the purpose of monitoring an experiment that classifies the stress state of a subject based on keystroke dynamics. This work is the first using the videos to detect stress through facial action units. Each video is approximately one hour long, though the actual length can vary by a few minutes depending on the typing speed of the subject. For the best knowledge of the authors, there is currently no other dataset publicly available that can be used for comparison.

A. Recording setup

Four different video streams were captured as part of the main experiment in [1]. Three of these streams captured only the subjects' typing on the keyboard, from the left, right, and above. The fourth stream consists of a frontal video of the subject's face during the experiment; this is the video data used in this work (see Fig. 1). The camera used was a Microsoft Life Studio Pro webcam with 1080p resolution at 30 frames per second. The frontal camera was positioned below the monitor which was used by the participants during the experiment.

B. Experimental Protocol

The face video data was collected as part of a larger experiment. The primary purpose of the larger experiment was to collect typing data from subjects in a neutral and stressed conditions to ascertain whether it was possible to discriminate between subjects' neutral and stressed typing. In addition to typing data, numerous ancillary data were collected such as blood pressure and ECG, to ascertain the stress state of the participants caused by the stressor.

Each subject in the experiment underwent the same protocol. In the beginning each subject was given a 30-minute rest period to place them in a neutral state. After the rest period, the subject was asked to provide a neutral typing sample, during which the first neutral face video was captured. The subject then performed a 15-minute stressor task, consisting of a multi-tasking exercise coupled with negative social evaluation from the experimenter. The subject then provided a stress typing sample followed by a second rest period; this 15-minute rest period returned the subject to a neutral state. A second neutral typing sample was then provided. Between each of the described phases, the subjects had a break of 2 min to they fill out questionnaires about their stress state. The overview of the phases recorded on video is shown in Fig. 2.

IV. Methodology

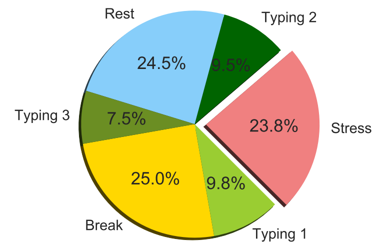
A. Features

In this work we focused on the use of a variety of facial Action Units (AUs) from the Facial Action Coding System (FACS) to distinguish stressed from non-stressed participants. Given the nature of the data only negative stress can be evaluated. To extract AUs from each frame of the videos, the toolbox OpenFace [25] was used which can detect 17 different AUs. The AUs used are shown in Table I. Most of the AUs are intensity values between 0-5, except AU7, AU23, and AU45 which are binary (present/absent). In Fig. 3 different intensities of the same AU are shown. In this work, we used for each frame only the AU values as feature (see Eq.1). To account for person-specific differences, the features were standardized per subject. Additionally, 2 minutes in between each phase were removed to avoid data from transitions.

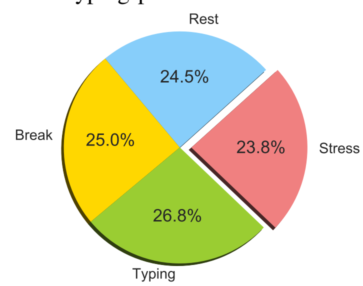
$$\mathbf{f} = (f_1 = \text{AU1}, f_2 = \text{AU2}, \ldots, f_{17} = \text{AU45}) \quad (1)$$

B. Data Visualization

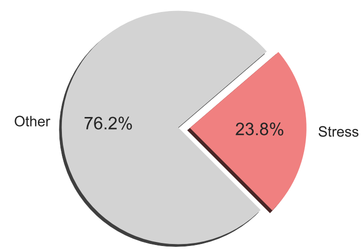
As an initial analysis, we visualized the behavior of each AU over time per subject. Each plot showed a different behavior of the AUs. Nevertheless, for each subject different single AUs were visibly correlated with the stress phase as is shown in Fig. 4. For this subject, AU20 which in [26] is also described as an indicator for fear is highly correlated to the stress phase. AU2, the outer brow raiser, is activated during the breaks. Also correlations with other phases were visible in other subjects. For that reason we used an unsupervised clustering for each subject. In Fig. 5 results of t-distributed Stochastic Neighbor Embedding (tSNE) [27] are shown for one subject. In Fig. 5a features recorded during different phases are colored differently, including the different typing phases. It is visible that the typing phase before and after the stressor are grouped separately. The final typing phase after resting, however, overlaps with the other two. In Fig. 5b the same data points are shown without distinguishing between the typing phases. It can be seen that differences between the phases exist. However, features during the resting phase overlap with features from the breaks in between main phases which is comprehensible. When distinguishing only between stress and non-stress phase, we can see that some features from the stress phase overlap the other phases (Fig. 5c).



(a) All classes shown by distinguishing the different typing phases.



(b) All typing phases joint together.



(c) Label distribution for binary problem, showing stress and grouping all remaining phases together.

Fig. 6: Label distribution for different classification problems.

TABLE II: Labels used for different classification problems.

2 Classes	4 Classes	6 Classed
1) Stress 2) All others	1) Break 2) Typing 1,2,3 3) Stress 4) Resting	1) Break 2) Typing 1 3) Stress 4) Typing 2 5) Resting 6) Typing 3

C. Classification

Given the clusters shown in Fig. 5, we defined three different classification problems (see Table II): 1) 6 class problem, distinguishing between all phases including each typing phase, 2) 4 class problem considering all typing phases as the same, and 3) binary classification problem distinguishing between stress phases and non-stress phase. For each classification problem, the label distribution varies (see Fig. 6). As the label distribution is only balanced in the 4 class problem, we used weighted accuracy on all classification results. Given that the behavior of the AUs was not visibly correlated between the five subjects, we trained one person independent model using LOSO method. We also trained 5 different models, one per subject, to evaluate person dependent classification using 5-fold cross-validation. For person independent classification the training set contained approximately 112,000 samples and testing 28,000 (total samples of one subject). During subject dependent classification 5-fold cross-validation was used, meaning that 80% of the data of one subject was used to train and the remaining 20% of the same subject was used for testing. Different simple classifiers were used: Random Forest, LDA, Gaussian Naive Bayes and Decision Tree.

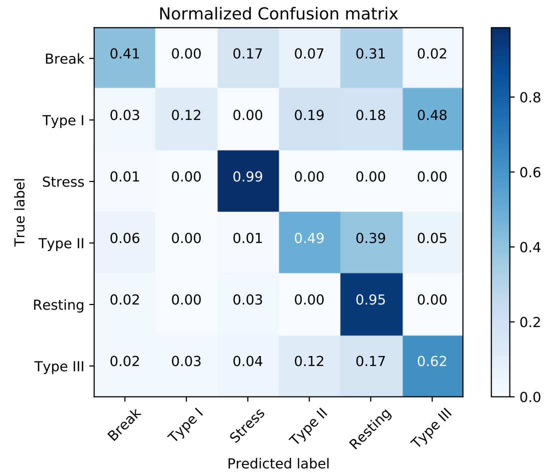
V. Results

The results obtained with each classifier for the different classification problems are shown in Table III. The results for subject independent classification vary from 29% for the 6 class problem to 74% for the binary classification. In contrast, person dependent results vary from 65% to 91% respectively.

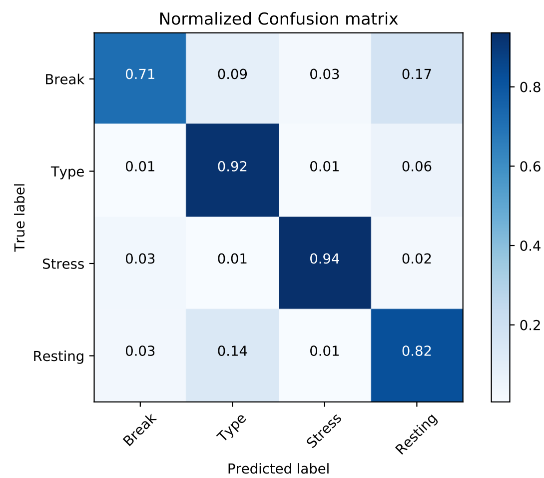
The best results for LOSO and subject-wise classification were obtained using Random Forest. Fig. 7 shows the average confusion matrix for the subject-wise classification using Random Forest.

VI. Conclusion

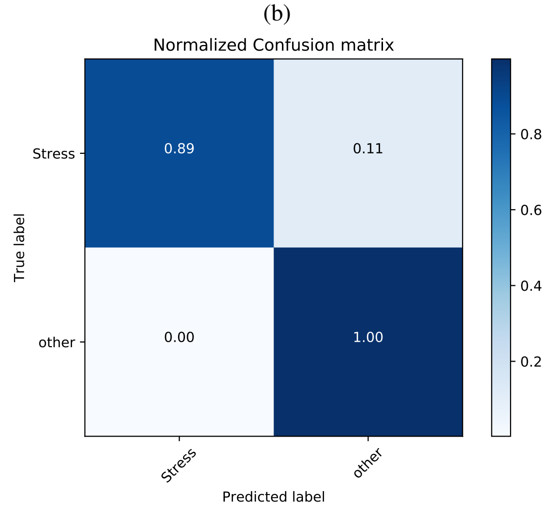
In this work, videos collected by Lau [1] were used for the first time to detect the stress state of five different participants using only facial AUs. Each participant was recorded while performing different tasks. By visualizing the features over time and through a tSNE plot, it was visible that AUs contain relevant information to distinguish the tasks. We formulated different classification problems to evaluate whether it was possible to correctly distinguish between each phase. As in previous work [23], subject independent classification achieved lower accuracy results than subject-specific classification. Given the small dataset used, this preliminary analysis



(a)



(b)



(c)

Fig. 7: Confusion matrices for subject-wise classification using Random Forest algorithm on a) 6 classes, b) 4 classes, and c) 2 classes.

TABLE III: Average accuracy results for person independent and dependent classification.

Person independent			
Classifier	2 Classes	4 Classes	6 Classes
Random Forest	0.75	0.49	0.41
LDA	0.74	0.47	0.33
Gaussian Naive Bayes	0.48	0.4	0.29
Decision Tree	0.68	0.34	0.29
Person dependent			
Classifier	2 Classes	4 Classes	6 Classes
Random Forest	0.93	0.83	0.83
LDA	0.89	0.74	0.65
Gaussian Naive Bayes	0.84	0.79	0.75
Decision Tree	0.89	0.74	0.67

showed that universal features were less accurate than person-specific features. However, more data and further analysis is needed to provide more conclusive results.

VII. Future Work

More data collected under the same recording conditions will be used in future work. In addition to facial AUs, other features will be analyzed. Also we intend to create a hybrid model that uses general features as basis, extended by person-specifc features. In order to compare this work with previous work [9], we will analyze how the detection of expressions for anger and fear would perform on this data to detect stress, compared to more diverse features.

References

[1] R. A. Maxion and S. Lau, "Facial expressions during stress and non-stress conditions: A benchmark video collection," Tech. Rep. CMU-CS-18-110, Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA 15213, 2018.

[2] J. Hassard, K. R. Teoh, G. Visockaite, P. Dewe, and T. Cox, "The cost of work-related stress to society: A systematic review.," *Journal of Occupational Health Psychology*, vol. 23, no. 1, p. 1, 2018.

[3] D. D. Ebert, F. Kählke, C. Buntrock, M. Berking, F. Smit, E. Heber, H. Baumeister, B. Funk, H. Riper, and D. Lehr, "A health economic outcome evaluation of an internet-based mobile-supported stress management intervention for employees.," *Scandinavian journal of work, environment & health*, vol. 44, no. 2, 2018.

[4] T. O. Davenport and J. Levin-Scherz, "Employers use stress diagnostics to define strategy and guide action," *Benefits Quarterly*, vol. 33, no. 1, p. 28, 2017.

[5] "Linkura program." https://www.linkura.com/linkuraprogrammet. Accessed: 2018-05-16.

[6] D. Carneiro, P. Novais, J. C. Augusto, and N. Payne, "New methods for stress assessment and monitoring at the workplace," *IEEE Transactions on Affective Computing*, 2017.

[7] B. Cvetković, M. Gjoreski, J. Šorn, P. Maslov, M. Kosiedowski, M. Bogdański, A. Stroiński, and M. Luštrek, "Real-time physical activity and mental stress management with a wristband and a smartphone," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp. 225–228, ACM, 2017.

[8] S. Clarke, L. G. Jaimes, and M. A. Labrador, "mstress: A mobile recommender system for just-in-time interventions for stress," in *Consumer Communications & Networking Conference (CCNC), 2017 14th IEEE Annual*, pp. 1–5, IEEE, 2017.

[9] J. S. Lerner, R. E. Dahl, A. R. Hariri, and S. E. Taylor, "Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses," *Biological psychiatry*, vol. 61, no. 2, pp. 253–260, 2007.

[10] H. Gao, A. Yüce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 5961–5965, IEEE, 2014.

[11] Y. Zhu, F. De la Torre, J. F. Cohn, and Y.-J. Zhang, "Dynamic cascades with bidirectional bootstrapping for action unit detection in spontaneous facial behavior," *IEEE transactions on affective computing*, vol. 2, no. 2, pp. 79–91, 2011.

[12] A. P. Association, "Understaunder chronic stress." https://www.apa.org/helpcenter/understanding-chronic-stress.aspx. Accessed: 2018-05-16.

[13] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[14] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *International Conference on Mobile Computing, Applications, and Services*, pp. 282–301, Springer, 2010.

[15] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *International Conference on Affective Computing and Intelligent Interaction*, pp. 125–134, Springer, 2011.

[16] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on galvanic skin response (gsr) controlled by zigbee," *Sensors*, vol. 12, no. 5, pp. 6075–6101, 2012.

[17] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*, pp. 209–214, IEEE, 2013.

[18] O. Simantiraki, G. Giannakakis, A. Pampouchidou, and M. Tsiknakis, "Stress detection from speech using spectral slope measurements," in *Pervasive Computing Paradigms for Mental Health*, pp. 41–50, Springer, 2016.

[19] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski, "Under pressure: sensing stress of computer users," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 51–60, ACM, 2014.

[20] M. Sysoev, A. Kos, and M. Pogačnik, "Noninvasive stress recognition considering the current activity," *Personal and Ubiquitous Computing*, vol. 19, no. 7, pp. 1045–1052, 2015.

[21] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from smartphones accelerometer data: a first step," *IEEE journal of biomedical and health informatics*, vol. 20, no. 4, pp. 1053–1060, 2016.

[22] E. Vildjiounaite, J. Kallio, J. Mäntyjärvi, V. Kyllönen, M. Lindholm, and G. Gimelfarb, "Unsupervised stress detection algorithm and experiments with real life data," in *Portuguese Conference on Artificial Intelligence*, pp. 95–107, Springer, 2017.

[23] A. Yüce, H. Gao, G. L. Cuendet, and J.-P. Thiran, "Action units and their cross-correlations for prediction of cognitive load during driving," *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 161–175, 2017.

[24] P. Ekman and W. V. Friesen, *Facial Action Coding System: Investigatoris Guide*. Consulting Psychologists Press, 1978.

[25] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pp. 1–10, IEEE, 2016.

[26] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 94–101, IEEE, 2010.

[27] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.