2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Person and Stressor Independent Generic Model for Stress Detection Using GSR

Dibyanshu Jaiswal, Debatri Chatterjee, Rahul Gavas, Ramesh Kumar Ramakrishnan, Arpan Pal
TCS Research, India

Abstract— **Stress detection is a widely researched topic and is important for overall well-being of an individual. Several approaches are used for prediction/classification of stress. Most of these approaches perform well for subject and activity specific scenarios as stress is highly subjective. So, it is difficult to create a generic model for stress prediction. Here, we have proposed an approach for creating a generic stress prediction model by utilizing knowledge from three different datasets. Proposed model has been validated using two open datasets as well as on a set of data collected in our lab. Results show that the proposed generic model performs well across studies conducted independently and hence can be used for monitoring stress in real life scenarios and to create mass-market stress prediction products.**

I. Introduction

Psychological stress is defined as a particular relationship between a person and the environment that is appraised by the person as taxing or exceeding his or her resources [1]. Psychological stress activates sympathetic nervous system of our body and results in bodily changes like increased blood pressure, heart rate etc. [2]. Thus, stress management is important to ensure better well-being. Hence, there is a requirement of a system which can monitor the stress levels continuously for a prolonged period. Stress responses can be measured using self-report questionnaires, behavioral changes or via physiological changes. First two measures require expert interventions and is not suitable for continuous or longitudinal measurement of stress in real world scenarios. Recent advances in non-invasive wearable bio-sensors enables continuous measurement of physiological responses like pulse rate, blood pressure, electrocardiogram (ECG), electro encephalogram (EEG), electromyogram (EMG), galvanic skin response (GSR), etc [3]. Such biofeedback based devices are portable, affordable and applicable for measuring stress in our everyday lives. Skin, the largest organ of human body, acts as an interface between our body and external environment. Sweat secretion changes the electrical property of skin. GSR measures subtle changes in electrical property of skin, such as, skin conductance level arising due to attention, emotional and other internal responses of autonomic nervous system. Wearable devices like Empatica E4, Shimmer3 GSR and RespiBAN professional are some of the devices used for recording the GSR signal.

Studies are being conducted for measurement of stress using GSR signal [4], [5] and a number of GSR based open source stress data sets are also available. Couple of such datasets are SWELL-KW [6] dataset generated from knowledge workers, Wearable Stress and Affect Detection (WESAD) [4] dataset. Diverse machine learning and deep learning based approaches are applied on these datasets for noise cleaning, GSR feature selection and classification of stress thereof. Stress is very subjective and the symptoms related to the nature of response varies across individuals for the same or similar stressors. In other words person and task(stressor) specific stress prediction models perform much better compared to a generic person and task independent model [4]. [7]. As a result, a classification model generated using one dataset often performs poorly on another dataset and hence it is difficult to create mass-market stress prediction products. In the present study, we propose a generic person and activity independent stress prediction model that gives comparable accuracy across datasets and individuals. The main contributions of our study are:

- identification of GSR features which discriminate stress, independent of *person* and *stressor*
- to create a person and activity independent generic model using above identified features for stress prediction
- application of the proposed model for predicting stress in real life scenarios and datasets

We have used two open datasets, called WESAD [4] and S-Test [5] along with our in-house collected stress dataset. Results show that the proposed model is a generic one and performs well across individuals belonging to different datasets with different stressors. Hence, the proposed model can be used to monitor stress in real life scenarios for various applications.

II. Related work

Stress is positively correlated with GSR [8]. Under stressful conditions, amount of sweating increases leading to increases the skin conductance level. Studies are being conducted on detection or classification of stressed vs non-stressed conditions using GSR signal. In [9], authors used GSR features and traditional machine learning approaches for detecting stress during arithmetic problem solving task under time pressure. In [10], authors proposed a GSR based driver stress detection system and reported a classification accuracy of 83%. In another GSR based study [11], authors analyzed relaxation response of an individual using various machine learning based approaches and concluded that decision tree based approach performs well. Majority of state of the art studies adopted multimodal fusion based approach for detection of stress. In [12], authors used speech and GSR signal for detecting stress. Other sensing modalities like heart rate [13], blood pressure [14], pupil diameter [15] are also being used in conjunction with GSR signal. However, most of these studies have detected stress under controlled or semi controlled laboratory environments. Literature suggests that for any physiological sensor based approach, there is a huge disparity in the accuracy reported using person and task specific stress prediction models and person/task independent models. In general, person/task specific models [16] , [17] achieved an excellent prediction accuracy but does not perform well on a completely unseen data (new person/task). In [17], authors monitored stress in daily work and found their person-specific ML models achieved a higher accuracy of 97% whereas the generic ones gives much lower accuracy of around 42%. Also in [16] authors achieved 90.0% accuracy when using a person-specific stress classification models but the same approach gives an accuracy of 58.8±11.6% for prediction of stress for new subjects. Thus, the replicability and reproducibility are two major issues that are faced by researchers as stress is highly dependent of one's genetics, coping abilities, and other factors like gender [18], response strategies etc. Hence, there is a need to evaluate the performance of a stress prediction model built using data from one study when used for prediction of stress from data collected for a completely different study with different participants and stressor. In this study, we propose an approach for creating a generic stress-prediction model and evaluate the performance of the model across studies conducted independently using different sets of participants and stressors.

III. Methodology

A. Stress dataset description

We have used three datasets to conduct the study. The first dataset - the WESAD dataset [4], includes physiological signals (GSR, ECG, EMG, respiratory signal and skin temperature) recorded using Empatica E4 [19] device from 15 participants under three different conditions (baseline, amusement and stress condition). Here Trier Social Stress Test (TSST) [20] protocol is used for inducing stress. It consists of five minutes of public speech, followed by a mental count down from 2023 by 17. In case of an error, the participants were requested to start over. State-Trait Anxiety Inventory (STAI) questionare was used to asses the ground truth of their current stress level. For our study, GSR data corresponding to baseline and stress conditions were used.

The second dataset - S-Test [5] uses Empatica device to record physiological signals from 21 participants. The stressor used here is Montreal Imaging Stress Task (MIST) [21]. After executing the task, participants reported their self-perceived stress level using Short STAI-Y anxiety questionnaires [22]. Initially, three levels of stress (low, medium and high) were induced using the MIST protocol, however, authors reported that they merged medium and high stress data because only two subjects reported high stress conditions. Physiological signals recorded during an initial baseline period is treated as no-stress data.

The third dataset ($DS3$) that we have used, was collected in-house from 15 participants following TSST protocol [23] using Shimmer3 [24]. After execution of the task, participants reported their self-perceived stress level using STAI questionnaires.

B. Proposed approach

Here, we present a study for stress detection on three different datasets. We experimented with dataset specific models and general models build using knowledge from all three data. Our proposed approach of creating the dataset independent general model is depicted in Fig. 1.

1) Normalization: To reduce the impact of inter-person variability while preserving the differences between the stress classes, we have performed normalization on each dataset. All datasets were converted to measure GSR in micro Simens and then re-sampled to a uniform sampling frequency of 4Hz. Next, GSR data corresponding to each participant was normalized using z-score normalization (by using its mean and standard deviation) strategy.

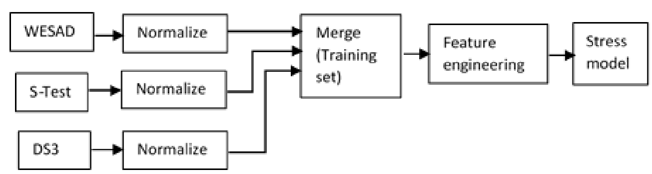

Fig. 1: Proposed approach of training utilizing knowledge from all three datasets

2) Feature engineering: A set of time and frequency domain GSR features are extracted on non-overlapping windows of duration 30 seconds. In total, 52 features were calculated from the normalized GSR signals. Among the 52, only 28 features were selected by using a in-house tool called *Feature Discovery Platform* (FDP) [25], which uses MICE and *Minimum Redundancy Maximum Relevance* (mRMR) approaches from a given super-set of features. The recommended features are listed in Table I. Such reduction in features can be performed using other commonly available approaches such as MICE or PCA also. However, FDP

TABLE I: Description of recommended GSR features

Feature Domain	Features
Time domain (18)	Min and max of tonic and phasic components respectively, skewness of phasic signal, var, std, kurtosis and skewness of tonic signal, no. of peaks, mean of phasic component, sum of positive derivatives, proportion of positive derivatives, first, second & third quantile of the signal, moment of GSR, median of all peak areas above threshold
Frequency domain (4)	relative tonic and phasic power, mean frequency, total power, median frequency.
Response features (6)	square average of phasic and tonic components, Accumulated GSR over a window, power of GSR signal spectrum in 2 bands from 0.15Hz to 0.4Hz, high freq band power ratio.

enables us to derive the optimal feature set (by maximizing the accuracy) automatically.

3) Building stress prediction model: The volume of training data have huge impact on the performance of any machine learning model. There are several approaches for increasing the volume of training data, however due to simplicity, we have used *Synthetic Minority Oversampling Technique* (SMOTE) [26] in a unique way. SMOTE is an oversampling approach used to handle class imbalance. It generates minority class samples at feature level, based on k-nearest neighbor of existing minority class samples. To increase the volume of data and variety within, we have artificially induced imbalance in the training dataset and generated new samples. The newly generated data instances were merged with original training instances to double the data volume and then the stress classifier is trained. Initially, we tried four classifiers, namely TreeBagger (RF), SVM, KNN and AdaBoost. However, we found that the best performance in terms of accuracy and F_{score} was achieved using random forest classifier. In order to optimize the training model, we varied the number of decision trees of the RF classifier and the model performance was judged through 10-fold cross validation on training data. In our case, the number of trees were chosen to be 100.

C. Stress prediction for completely unseen data

In order to create a person and activity independent test set, we adopted two approaches *Scheme I* - we selected 10% of the subjects, randomly from each dataset and created a new test set. Remaining 90% of the subjects from each dataset were merged together to form the training set. Number of participants selected from each dataset are shown in Table II. Thus, our test set contains physiological responses of participants belonging to different demographics and collected during different activities. Our stress prediction model is built on GSR data of remaining 44 participants across 3 datasets. Normalized GSR recording corresponding to no-stress and stress conditions were used to form the training set. *Scheme II* - Similar to Scheme I, with a train test spilt of 80-20%. Thus, in this case number of unseen test participants were 13 and training model was build using data from remaining 38 participants.

During testing, physiological sensor data of test participants are first z-score normalized, and further divided into windows of duration 30 seconds. Recommended 28 signal property based features are extracted on each of these windows. Finally, these features are fed to the trained classifier to get the final predictions (i.e stress/no-stress). The performance of our training model has been evaluated in terms of accuracy, sensitivity, specificity and F_{score}.

IV. Results and Discussion

Inter-dataset differences in stress prediction and comparison with SOA: The most discriminating 28 features (Table I), we created separate classification models for each of the 3 datasets. The average classification performance for various datasets using our selected feature set is shown in Table III, w.r.t GSR alone and using multiple sensors. We achieved very high accuracy and F_{score} for stress prediction for all the datasets when using same validation approach as used in SOA (leave one subject out for [4], [5] and leave one sample out for [23]). For WESAD, authors reported [4] an average accuracy of 79% (F_{score} = 0.75) for stress vs. no-stress classification using GSR signal only, and 87% using all wrist-worn (GSR, Temp, Acc and BVP) sensor data. However, in proposed approach with GSR signal only, we obtained an average accuracy of 88% with an F_{score} of 0.82. For S-Test dataset, authors [5] reported an accuracy of approximately 58% using GSR signal only and 73% with all sensors and an average F_{score} of 0.47. Whereas, our proposed method achieves an accuracy of 73% and F_{score} of 0.79. Lastly, for DS3 dataset, authors reported an F_{score} of 0.64 while using GSR signal only. With our proposed GSR features, we obtained an F_{score} of 0.66. Thus, we can conclude that, a) The proposed 28 features performs better than the SOAs in classifying stress vs. no-stress for each of the datasets b) better accuracy than the SOA, enables its use as the benchmark for prediction comparisons

TABLE II: Praticipant Counts for Train and Test

Scheme	Dataset	Total subjects	Train subjects	Test subjects
I	WESAD	15	13	2
	S-Test	21	18	3
	DS3	15	13	2
II	WESAD	15	11	4
	S-Test	21	16	5
	DS3	15	11	4

TABLE III: Comparison of Classification results (using LOSO approach), using proposed features and SOA

Dataset	Sensor used	SOA reported		Proposed approach	
		Acc (%)	Fscore	Acc (%)	Fscore
WESAD	GSR	79	0.75	**88**	**0.82**
	Multiple	87	0.84		
S-Test	GSR	58	–	**73**	**0.79**
	Multiple	73	0.47		
DS3	GSR	–	0.64	**66.3**	**0.66**

Next, we analyse the prediction performance across datasets. Here, our proposed stress prediction model created using one dataset is used to predict stress from another dataset. The average classification performance in terms of accuracy and F_{score} are presented in Table IV. It is observed that except the cases where WESAD has been used as test dataset, prediction performance is very low and inconsistent. Corresponding sensitivity and specificity values are also very low. The cases, where WESAD is used as the test set, the accuracy is comparatively better but still less than the case where the classification model is trained on WESAD itself (Table III). Thus, we can conclude that though dataset (i.e activity and subject specific) specific stress prediction model performs well for predicting stress against the same dataset, it

TABLE IV: Stress prediction performance using proposed model across datasets

Training set	Test set	Accuracy (%)	Sens.	Spec.	F_{score}
S-Test	DS3	49	0.72	0.37	0.50
WESAD	DS3	58	0.53	0.6	0.48
DS3	WESAD	75	0.82	0.71	0.70
S-Test	WESAD	80	0.91	0.74	0.77
WESAD	S-Test	54	0.38	0.95	0.55
DS3	S-Test	56	0.5	0.71	0.62

TABLE V: Classification accuracy for proposed generic model using scheme I, scheme II and 10 fold cross-validation

Scheme	No of features	Acc. (%)	Sens.	Spec.	F score
I	28	86	0.87	0.85	0.82
II	28	78	0.76	0.79	0.7
10FCV on full set	28	82	0.76	0.87	0.80

gives inconsistent and lower accuracy when tested on unseen user data from another dataset.

Stress prediction using proposed generic model: Finally, we analyzed the performance of our proposed approach of a generic model. The average stress prediction accuracy and F_{score} across test participants is presented in Table V. As explained in Section III-C, the test set was constructed using two schemes. It is observed that our proposed model is performing well with an accuracy of 86% and F_{score} of 0.82 for scheme I, irrespective of the demography and stressor. For scheme II, the accuracy achieved was 78% with an F_{score} of 0.7. We have also reported the *10 fold cross-validation* (10FCV) accuracy on the combined dataset using proposed features and approach. All the three evaluation cases show consistent results over different metrics. Hence, the proposed model is a generic stress prediction model that performs well across datasets with different demographics/tasks and also outperforms state of the art approaches.

V. Conclusions and Future Scope

In the present work we have proposed a generic stress prediction model which is person and stressor agnostic. The proposed approach recommends a set of features which enables the development of a generic stress classifier. The stress prediction models built on each dataset using those features outperforms the corresponding state of the art models. Moreover, the proposed generic model has very high accuracy (86%) and F_{score} for a test set created using a combination of unseen participants with different stressors. Thus, the proposed model is generic in nature that can be used successfully across participants and activities for monitoring stress in real time. In future, we would like to validate our findings on diverse datasets with and without transfer learning. We have used SMOTE here to augment the training data. In future we would also like to apply Generative adversarial network (GAN) based approaches for the same.

References

[1] Richard S Lazarus and Susan Folkman, *Stress, appraisal, and coping*, Springer publishing company, 1984.

[2] ANDREW STEPTOE and LENA BRYDON, "Psychosocial factors and coronary heart disease: the role of psychoneuroimmunological processes," in *Psychoneuroimmunology*. Elsevier, 2007.

[3] Bin Yu, Mathias Funk, Jun Hu, Qi Wang, and Loe Feijs, "Biofeedback for everyday stress management: A systematic review," *Frontiers in ICT*, vol. 5, pp. 23, 2018.

[4] Philip Schmidt and et.al., "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *ACM international conference on multimodal interaction*, 2018.

[5] Martin Gjoreski and et.al., "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, 2017.

[6] Saskia Koldijk and et.al., "The swell knowledge work dataset for stress and user modeling research," in *International conference on multimodal interaction*, 2014.

[7] Jonghwa Kim and Elisabeth André, "Emotion recognition based on physiological changes in music listening," *IEEE Tran. on pattern analysis and machine intelligence*, vol. 30, no. 12, 2008.

[8] Jorn Bakker and et. al., "What's your current stress level? detection of stress patterns from gsr sensor data," in *IEEE 11th international conference on data mining workshops*, 2011.

[9] Cornelia Setz and et.al., "Discriminating stress from cognitive load using a wearable eda device," *IEEE Tran. on information technology in biomedicine*, vol. 14, no. 2, 2009.

[10] N El Haouij and et.al., "Feature extraction and selection of electrodermal reaction towards stress level recognition: Two real-world driving experiences," *47emes Journees de Statistique de la SFdS (JdS)*, 2015.

[11] Raquel Martinez and et.al., "A self-paced relaxation response detection system based on galvanic skin response analysis," *IEEE Access*, vol. 7, 2019.

[12] Hindra andet.al. Kurniawan, "Stress detection from speech and galvanic skin response signals," in *IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2013.

[13] Xu Xu, *Analysis on mental stress/workload using heart rate variability and galvanic skin response during design process*, Ph.D. thesis, Concordia University, 2014.

[14] Atlee Fernandes and et.al., "Determination of stress using blood pressure and galvanic skin response," in *International Conference on Communication and Network Technologies*. IEEE, 2014.

[15] Jing Zhai and et.al., "Realization of stress detection using psychophysiological signals for improvement of human-computer interactions," in *IEEE SoutheastCon*. IEEE, 2005.

[16] Saskia Koldijk and et.al., "Detecting work stress in offices by combining unobtrusive sensors," *IEEE Tran. on affective computing*, vol. 9, no. 2, 2016.

[17] Yoshiki Nakashima and et.al., "Stress recognition in daily work," in *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, 2015.

[18] Alexandros Liapis and et.al., "Stress recognition in human-computer interaction using physiological and self-reported data: a study of gender differences," in *Panhellenic Conference on Informatics*, 2015.

[19] "Empatica E4 Device," `https://www.empatica.com/research/e4/`, [Online; accessed 20-July-2021].

[20] Clemens Kirschbaum and et.al., "The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, 1993.

[21] Katarina Dedovic and et.al., "The montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain," *Journal of Psychiatry and Neuroscience*, vol. 30, no. 5, 2005.

[22] Gerald Bauer and et.al., "Can smartphones detect stress-related changes in the behaviour of individuals?," in *IEEE PerCOM*, 2012.

[23] Rahul D et.al. Gavas, "A sensor-enabled digital trier social stress test in an enterprise context," in *IEEE EMBC*, 2019.

[24] "Shimmer3 Device," `https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor`, [Online; accessed 20-July-2021].

[25] Snehasis et. al. Banerjee, "Towards wide learning: Experiments in healthcare," *arXiv preprint arXiv:1612.05730*, 2016.

[26] Nitesh V Chawla and et. al., "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, 2002.