pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210607

Continuous Stress Detection of Hospital Staff Using Smartwatch Sensors and Classifier Ensemble

Muhammad Ali FAUZI[1] and Bian YANG
Norwegian University of Science and Technology, Gjøvik, Norway

Abstract. High stress levels among hospital workers could be harmful to both workers and the institution. Enabling the workers to monitor their stress level has many advantages. Knowing their own stress level can help them to stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. This monitoring task can be enabled by using wearable devices to measure physiological responses related to stress. In this work, we propose a smartwatch sensors based continuous stress detection method using some individual classifiers and classifier ensembles. The experiment results show that all of the classifiers work quite well to detect stress with an accuracy of more than 70%. The results also show that the ensemble method obtained higher accuracy and F1-measure compared to all of the individual classifiers. The best accuracy was obtained by the ensemble with soft voting strategy (ES) with 87.10% while the hard voting strategy (EH) achieved the best F1-measure with 77.45%.

Keywords. Stress detection, Hospital, Machine Learning, Ensemble, Smartwatch

1. Introduction

Over recent years, stress has become an interesting topic in today's hectic world. There has been increasing awareness in many countries about the rise of work-related stress. Hospital is possibly one of the most important workplaces to be alarmed about this issue. Many studies reported that many hospital workers suffer from work-related stress [1,2]. Although stress at some level is normal, chronic stress can harm our physical, mental, and emotional wellbeing [3,4]. Specifically for hospital, many studies suggested that higher stress level has a relationship with low patient safety [5,6]. Another study also reported that higher stress level is significantly correlated with riskier cybersecurity practices [7].

Monitoring hospital workers' stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly [8]. Besides, this monitoring can help for early diagnosis of mental illness and disorders. The most common way to assess stress level is by using questionnaires (e.g., Perceived Stress Scale [9], Perceived Stress Questionnaire [10], etc.). However, this method takes time so that it is not convenient to be performed every day for continuous monitoring.

[1] Corresponding Author: Muhammad Ali Fauzi, Norwegian University of Science and Technology, Teknologivegen 22, 2815 Gjøvik, Norway; E-mail: muhammad.a.fauzi@ntnu.no.

The other stress level assessment method is by measuring the physiological responses related to stress such as heart rate, blood pressure, skin conductance, respiration activity, etc. Some sensors can be used to conduct the measurement task (e.g., electrocardiogram (ECG) to measure the heart rate, galvanic skin response (GSR) for skin conductance, etc.). The recent advance in wearable devices with sophisticated built-in sensors makes it feasible to passively collect multimodal data from people's daily lives for automatic continuous stress detection purposes. However, some wearable devices have a very low usability and not convenient to wear during work (e.g., chest-worn devices, finger placed GSR sensors, etc.) [11].

Smartwatch has recently emerged as a new platform that provides many successful applications. These devices have several built-in sensors that are useful for stress monitoring including Blood Volume Pulse (BVP), Electrodermal Activity (EDA), temperature, accelerometer, etc. Besides, the use of watches is well known and has a high degree of social acceptance by their ubiquity in everyday life [12]. Therefore, it has a high potential to be applied for multi-modal-based continuous stress detection.

Many previous works have been successfully leveraging multi-modal sensors data and machine learning methods to build automatic stress detection. The popular machine learning methods used are Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression [13,14,15,16]. In this work, we propose a multi-modal based continuous stress detection method using classifier ensemble and give comparative analysis between individual classifiers. Classifiers ensemble is a set of base classifiers whose individual classification outputs are combined in some way in order to enhance classification accuracy [17]. The individual classifiers used for this works include Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT).

2. Proposed Work

2.1. Dataset

This research is based on the WESAD [13] dataset, which is available to the public. It includes data from 15 people who were measured with the Empatica E4 wrist-worn device and chest-worn RespiBAN device. However, because the focus of this work is on smartwatch sensors, only E4 data is used in this analysis. The E4 gadget incorporates skin temperature (ST), accelerometers (ACC), electrodermal activity (EDA), and blood volume pulse sensors (BVP) sensors. Data from three separate affective states (stress, amusement, and relaxation) were obtained during the data collection process. The stress situation lasted about 10 minutes, the amused situation 6.5 minutes, and the relaxed situation 20 minutes. For the stress detection task in this study, the amusement and relaxation classes were merged into one class: non-stress. As a result, the problem under investigation was binary (stress and non-stress).

2.2. Features

In this study, we used the data from all of the sensors available in the smartwatch including ACC, EDA, ST, and BVP. To extract the features, a sliding window with a window shift of 0.25 seconds was used to segment the data. Furthermore, the ACC features were computed with a five-second window size, as this is a common window length for acceleration-based context detection [18]. Meanwhile, all other physiological features were calculated with a window size of 60 seconds following the suggestion by Kreibig et al. [19]. The AC, EDA, and ST features were extracted based on prior work by [20]. The features extracted including some statistical features (mean, standard deviation, maximum, and minimum). Besides, some derivatives and Discrete Wavelet Transform (DWT) were also applied to the data to extract other statistical features. Meanwhile, for BVP, statistical features (mean, standard deviation) were also computed. Moreover, some features based on energy in different frequency bands were also calculated.

2.3. Classifier

Seven machine learning methods were used as classifiers for stress detection tasks including Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT). In addition, we also used two ensemble methods. In order to do stress detection, the ensemble technique trains numerous classification methods and then combines them using particular approach [21]. It is important to take note that the performance of the ensemble methods cannot be guaranteed to be higher than the best individual method in the ensemble. However, it would significantly minimize the chances of picking a poor-performing classifier [17].

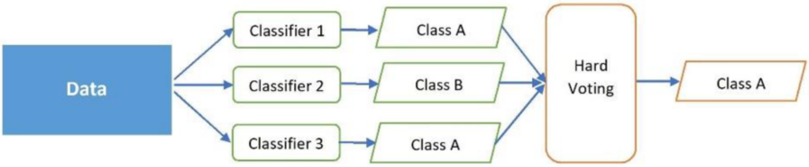

Figure 1. The hard voting strategy.

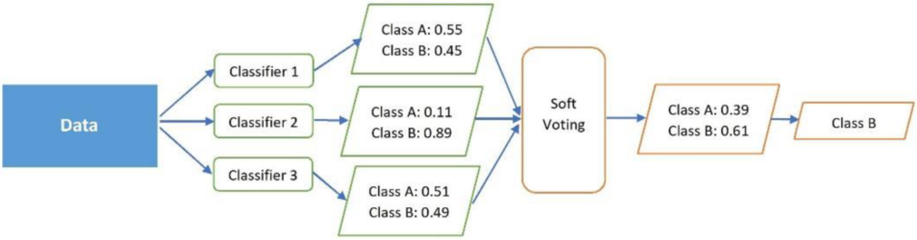

Figure 2. The soft voting strategy.

In this study, we employed three classification methods to build the ensemble learning method. Three individual classifiers with the highest accuracy were selected for the ensemble. Two ensemble strategies were used in this work as follows:

1. Hard voting (hard): As depicted in Figure 1, each classifier had one vote, and the class of the data was determined by the majority vote.

2. Soft voting (soft): As depicted in Figure 2, each classifier calculated the probability of each class in the first step. Then, the probabilities of each class from all classifiers were averaged, and the final class of the data was the one with the greatest average probability value.

2.4. Performance Evaluation

All classifiers were tested using the leave-one-subject-out (LOSO) cross-validation (CV) approach, which shows how a model will generalize and perform on previously unseen data. Several measurements including Accuracy, Precision, Recall, and F1-measure were employed for classifier performance evaluation.

3. Result

The stress detection result using individual classifiers is shown in Table 1 while the result using classifiers ensemble is displayed in Table 2. Table 1 depicts that all classifiers work quite well to conduct a stress detection task. All of them show adequate performance with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% following by LR with a slight difference (85.46%). At third place was NN that has a slight margin to the first and second place (84.76%). These three top classifiers were then used for the ensemble methods. Meanwhile, the lowest accuracy was achieved by KNN with a value of only 73%.

In terms of precision, LR has the best precision among other individual classifiers with a value of 77.53%. Furthermore, in terms of recall, RF has the highest value with 89.87%. However, the precision of RF is quite low (69.17%) so that it could not obtain the highest F1-measure. It means that RF tends to successfully detect almost all of the stress data available but many non-stress data are incorrectly labeled as stress. Meanwhile, LR has a more balance precision and recall so that it could achieve the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1-measure (52.43%).

The ensemble methods were build using the three best individual classifiers from the previous results (RF, LR, and NN). Table 2 shows that both ensemble methods obtained higher performance compared to all of the individual classifiers. Generally, most individual classifiers have their own inherent defects [22] and their performance is also domain-dependent [23]. By combining some classifiers, the advantage of one classifier is expected to cover the shortcomings of other classifiers so that the performance can be improved. Soft and hard voting have different strategies to combine the result from the individual classifiers so that they can lead to different decisions.

The result displayed in Table 2 shows that ES (soft voting) has a higher accuracy than EH (hard voting). In contrast, EH has a better performance in terms of F1-measure. Generally, the soft voting strategy tends to get better performance than the hard voting strategy as it takes into account more information. Soft voting is smoother as it uses probability information to get the final decision. However, the additional information could also lead to a worse decision. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

Table 1. Stress Detection Result Using Individual Classifiers (%)

Method	Accuracy	Precision	Recall	F1-measure
NB	79.26	57.58	73.31	60.67
SVM	84.60	76.29	80.51	75.01
NN	**84.76**	76.53	80.12	74.97
KNN	73.71	52.31	63.74	52.43
LR	**85.46**	**77.53**	82.16	**76.25**
RF	**86.61**	69.17	**89.87**	73.05
DT	79.25	66.90	73.59	66.81

Table 2. Stress Detection Result Using Classifiers Ensemble (%)

Method	Accuracy	Precision	Recall	F1-measure
EH	86.99	76.00	88.02	**77.45**
ES	**87.10**	76.11	86.75	75.91

4. Conclusion

Enabling the workers to monitor their own stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. This monitoring task can be enabled by using wearable devices to measure related physiological responses. Smartwatch is one of the devices that can be used for this task due to its usability for the working environment and its built-in sensors. In this work, we propose a multi-modal based continuous stress detection method using some individual classifiers and classifier ensembles.

The experiment results show that all classifiers work quite well to detect stress with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% while KNN has the lowest accuracy with 73%. In terms of F1-measure, LR achieved the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1measure (52.43%). The results also show that the ensemble method obtained higher performance compared to all individual classifiers. In this study, the advantage of one classifier can cover the shortcomings of other classifiers so that the accuracy can be improved. Furthermore, the results also show that ES (soft voting) has higher accuracy than EH (hard voting) in this study but EH has a better F1-measure than ES. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

Our experimental study for the effect classifier ensemble is limited by the WESAD dataset that uses only two classes: stress and non-stress. In future work, the effect of the use of the ensemble method can be tested on a dataset that provides different stress levels (e.g., low stress, moderate stress, and high stress). Besides, a new dataset with more subjects could be created in the future in order to test the reliability of the proposed methods. The future dataset could also include not only label based on the intervention like in the WESAD dataset, but also the label from user-filled questionnaires (e.g., PSS).

References

[1] Marine A, Ruotsalainen JH, Serra C, Verbeek JH. Preventing occupational stress in healthcare workers. Cochrane Database of Systematic Reviews. 2006;(4).

[2] Weinberg A, Creed F. Stress and psychiatric disorder in healthcare professionals and hospital staff. the Lancet. 2000;355(9203):533–537.

[3] Pickering TG. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. Current hypertension reports. 2001;3(3):249–254.

[4] Wang Y, Chen R, Zhang L. Reliability and validity of generalized anxiety scale-7 in inpatients in Chinese general hospital. J Clin Psychiatr. 2018;28:168–71.

[5] Welp A, Meier LL, Manser T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. Frontiers in psychology. 2015;5:1573.

[6] Tsiga E, Panagopoulou E, Montgomery A. Examining the link between burnout and medical error: A checklist approach. Burnout Research. 2017;6:1–8.

[7] Fauzi MA, Yeng P, Yang B, Rachmayani D. Examining the Link Between Stress Level and Cybersecurity Practices of Hospital Staff in Indonesia. In: The 16th International Conference on Availability, Reliability and Security; 2021. p. 1–8.

[8] Liao W, Zhang W, Zhu Z, Ji Q. A real-time human stress monitoring system using dynamic Bayesian network. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops. IEEE; 2005. p. 70–70.

[9] Cohen S, Kamarck T, Mermelstein R. Perceived stress scale (PSS). J Health Soc Beh. 1983;24:285.

[10] Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, et al. Development of the Perceived Stress Questionnaire: a new tool for psychosomatic research. Journal of psychosomatic research. 1993;37(1):19–32.

[11] Spagnolli A, Guardigli E, Orso V, Varotto A, Gamberini L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In: International Workshop on Symbiotic Interaction. Springer; 2015. p. 87–98.

[12] Lazaro MJS, Lim J, Kim SH, Yun MH. Wearable technologies: acceptance model for smartwatch adoption among older adults. In: International Conference on Human-Computer Interaction. Springer; 2020. p. 303–315.

[13] Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM international conference on multimodal interaction; 2018. p. 400–408.

[14] Garg P, Santhosh J, Dengel A, Ishimaru S. Stress Detection by Machine Learning and Wearable Sensors. In: 26th International Conference on Intelligent User Interfaces; 2021. p. 43–45.

[15] Indikawati FI, Winiarti S. Stress detection from multimodal wearable sensor data. In: IOP Conference Series: Materials Science and Engineering. vol. 771. IOP Publishing; 2020. p. 012028.

[16] Siirtola P. Continuous stress detection using the sensors of commercial smartwatch. In: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; 2019. p. 1198–1201.

[17] Fauzi MA, Yuniarti A. Ensemble method for indonesian twitter hate speech detection. Indonesian Journal of Electrical Engineering and Computer Science. 2018;11(1):294–299.

[18] Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th international symposium on wearable computers. IEEE; 2012. p. 108–109.

[19] Kreibig SD. Autonomic nervous system activity in emotion: A review. Biological psychology. 2010;84(3):394–421.

[20] Zhang Y, Haghdan M, Xu KS. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers; 2017. p. 54–57.

[21] Fauzi MA, Bours P. Ensemble Method for Sexual Predators Identification in Online Chats. In: 2020 8th International Workshop on Biometrics and Forensics (IWBF). IEEE; 2020. p. 1–6.

[22] Li L, Zhang Y, Zou L, Li C, Yu B, Zheng X, et al. An ensemble classifier for eukaryotic protein subcellular location prediction using gene ontology categories and amino acid hydrophobicity. PLoS One. 2012;7(1):e31057.

[23] Kumar BS, Ravi V. Text document classification with pca and one-class svm. In: Proceedings of the 5th international conference on frontiers in intelligent computing: theory and applications. Springer; 2017. p. 107–115.