4. Interdisciplinary Conference on Electrics and Computer (INTCEC 2024)
11-13 June 2024, Chicago-USA

Personalized Stress Detection from Chest-Worn Sensors by Leveraging Machine Learning

Md Nahid Hasan
Computational and Data Science
Middle Tennessee State University
Murfreesboro, TN, USA
mh2ay@mtmail.mtsu.edu

Mithun Saha
Computer Science
University of Memphis
Memphis, TN, USA
msaha1@memphis.edu

Md Sultan Mahmud
Electrical and Computer Engineering
University of Memphis
Internal Medicine
University of Tennessee Health Science Center
Memphis, TN, USA
mmahmud@memphis.edu, mmahmu10@uthsc.edu

Samir Poudel
Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
sp2ai@mtmail.mtsu.edu

Satish Wagle
Computational and Data Science
Middle Tennessee State University
Murfreesboro, TN, USA
sw8k@mtmail.mtsu.edu

Khem Poudel
Computational and Data Science
Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
khem.poudel@mtsu.edu

***Abstract*—Stress detection is crucial for individual well-being. We investigated stress detection using machine learning (ML) from chest-worn sensor data. We used the WESAD dataset that recorded physiological data via a chest-worn sensor when 15 subjects were in stress and non-stress. We empirically evaluated different combinations of window sizes and ML models for stress detection. Our analysis shows XGBoost achieved the highest accuracy of 99.71%, F1-score 99.57%, and area under the curve (AUC) 99.61% in detecting population-level stress with an optimal window size of 500 ms. We also developed personalized stress detection models using the same window size. Subject-level results show stress can be more accurately detected (accuracy 99.95%, F1-score 99.90%, and AUC 99.87%) with personalized models. Our findings also demonstrate a 2% improvement in stress detection compared to prior studies. This work could prove beneficial for quick stress detection and understanding the cause of stress.**

***Index Terms*—Stress detection, Wearable sensors, Machine learning, Personalization, XGBoost.**

I. Introduction

Stress is characterized by the activation of the body's 'fight' or 'flight' response to challenging situations caused by internal and external factors needing urgent attention or action. Everyone experiences stress in their daily lives. When a stressful event occurs, the autonomic nervous system, which comprises the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) is activated to prepare the body. SNS elevates physiology by increasing heart rate, while PNS in an effort to restore homeostasis, brings it back to resting state as the threat or challenge diminishes. But, albeit necessary, frequent activation of the autonomic nervous system due to repeated and prolonged exposure to stress has detrimental effects on mental, physical, financial, and social well-being [1], [2]. Stress has been linked to life-threatening physical conditions and diseases such as obesity [3], diabetes, coronary heart disease, and even cancer [4]. Moreover, it can contribute to mental illness [5], and disrupted sleep patterns [6] in individuals, with unhealthy coping strategies such as overeating [7], alcoholism, drug abuse [8], or social isolation [9] being the undesired result.

Given all the harmful consequences, efforts have been ongoing to help individuals deal with stress in various ways. A recent work [10] has shown that prompting participants to log stressors for detected stressful events and providing reflective interventions through weekly visualizations has the potential to result in study-long reduction of stress in individuals. Conversely, a more active and timely intervention strategy like Ecological Momentary Intervention (EMI) addresses stress in the moment [11]. Among other factors, the effectiveness of EMIs largely depends on being able to assess the vulnerable state in question accurately. Irrespective of the modality of intervention, an increased ability to correctly identify the majority of these stressful moments proximal to their occurrences can enhance their effectiveness in reducing stress and promoting overall wellness.

Ubiquitous and unobtrusive stress tracking in daily life is thus essential for gaining comprehensive and reliable insights into stress patterns. Stress can be detected using data collected from various sources. For example, [12] detected the stress of drivers by analyzing facial expressions using a near-infrared (NIR) camera. To separate the stressful states of a computer user, [13] used features computed from pupil dilation and periorbital temperature. Widespread use of smartphones prompted [14] to derive keyboard typing patterns using smartphone accelerometers and gyroscope sensor data for detecting the stress of a smartphone user. Vhaduri et al. [15] demonstrated that stress can be detected solely from location transitions using GPS traces. With the growing popularity and use of different social media platforms, natural language-based stress detection [16] has the potential to provide insights into mental disorders. Although useful, the above modalities of stress detection suffer from their inability to continuously monitor individuals due to infrequent interactions, inconvenience, privacy concerns, or cost of data collection.

The rapid development of wearable devices over the past decade has facilitated more continuous monitoring of individuals for physiological data collection in real-world settings. Physiological signals are internal to the human body, representing outcomes of different biological processes that produce measurable changes in electrical, mechanical, or chemical activity. Some widespread internal signals include electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR), respiration (RSP), and photoplethysmography (PPG). These signals can be utilized to produce measurable metrics of human physiology in the form of biomarkers like heart rate (HR) [17], heart rate variability (HRV) [18], respiratory rate (RR) [19], blood pressure (BP) [20], oxygen saturation (SpO2) [21], skin conductance level (SCL) [22] among others. It has been shown that these biomarkers [23]–[29] exhibit quantifiable changes during stress exposure that can be utilized to build deep learning and machine learning (ML) models for automatic detection of stress.

Significant advancement and evolution of deep learning in the last decade have enabled stress detection works to achieve impressive accuracy of more than 90% using lab study protocols. To address the challenges of the current mental healthcare system, [30] developed an end-to-end deep learning framework for stress detection requiring less human involvement with a reported accuracy of 91.52%. Gil-Martin et al. employed three signal processing techniques on all the available physiological signals of participants before supplying them to a CNN-based deep learning architecture for stress detection [31]. They report an average accuracy of 96.62% and an F1-score of 96.63% in a leave one subject out evaluation setup. Instead of raw, unprocessed data, [32] transformed raw data from the chest band into images for input to a deep model, achieving an accuracy of 94.8%. Likewise, many other works have also applied deep learning methods for stress detection. Still, deep learning has disadvantages, especially with overfitting, enormous amounts of data, computational resources, and interpretability. When these factors become crucial, ML-based stress detection can offer a potential solution. To run stress detection on smartphones and smartwatches, [33] extracted simple mean and min based features from EDA and SCR to build a ML model with an accuracy of 92% across all subjects. Similarly, different sensor signals were utilized by [34] with other ML models achieving an F1-score of 83.34% for binary stress classification. For continuous stress detection, Siirtola [35] found LDA gives an average accuracy of 87.4% across participants when trained with features from ST, BVP, and HR using a sliding window of 120 seconds. However, most of these works only reported results of a population-based model, whereas stress detection generalizability is problematic due to between-subject and between-context variability as [35] reported a significant variation in model performance among participants. In this work, we investigate the potential of personalized stress detection using ML models with minimal features, excluding resource-intensive signal processing for computational efficiency. In addition, we also explore the performance of population-based stress detection.

To the best of our knowledge, prior works demonstrated stress classification using ML utilizing signals from different sensors with decent results. However, compared to recent deep learning powered stress detection, their performance is relatively lower.

Our main contributions are as follows:

- We empirically evaluated the optimal window size that can identify the stress vs. non-stress states with higher accuracy.
- We explored stress detection with six features (minimum, maximum, mean, standard deviation, skewness and kurtosis) that improved model performance by 2% at the population level.
- We also investigated the potential of personalized stress detection. Majority of the participants registered a stress classification accuracy of 99%, with the lowest accuracy achieved was 98% for a single subject.

The rest of the paper is organized as follows. We describe the methodology in Section II. Subsequently, Section III presents the results, and Section IV concludes the paper with the key findings.

II. Materials and Methods

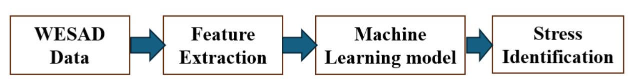

Fig. 1. Workflow diagram Stress vs non stress detection

A. Dataset Description

Stress detection has greatly benefited from some high-quality publicly available datasets. Wearable stress and affect detection (WESAD) [36] is one such dataset with a wide range of physiological signals collected via chest and wrist devices. Hence, to demonstrate the effectiveness of the proposed stress detection approach, we used this dataset that consists of 15 participants (12 males and 3 females). Participants wore a RepsiBAN Professional on the chest and an Empatica E4 on the non-dominant wrist while stress and non-stress stimuli were presented. The chest-based device with a sampling rate of 700 Hz, recorded body motion along with RESP, ECG, EMG, EDA, and TEMP, while the wrist-based device recorded BVP, EDA, TEMP, and ACC at varying sampling rates.

In order to induce different affective states, three primary sessions were conducted–Baseline, Stress, and Amusement. For the purpose of de-exciting participants following both stress and amusement sessions, a meditation session was held with an additional ten-minute rest period preceding right after the stress session. During the baseline session, participants read magazines for about 20 minutes, sitting/standing at a table to maintain a neutral position. The stress session lasted approximately 10 minutes and was divided into two phases of equal duration to administer a well-validated Trier Social Stress Test [37]. The first phase involved public speaking, in which participants had to talk about their strengths and weaknesses in front of a panel. In the second phase, they had to perform a mental arithmetic exercise of counting down from 2023 to zero in increments of 17, with any error resulting in a restart. In the amusement session, participants watched 11 funny videos totaling 6.5 minutes sourced from a corpus [38] and the authors' list of preferences. The participants provided self-reports after each session, including rest and meditation, which further enriched the dataset.

B. *Feature Extraction*

We experimented with different window sizes (e.g., 100 ms, 200 ms, 300 ms, 400 ms, 500 ms, 600 ms, 700 ms, 800 ms, 900 ms and 1000 ms) to find the optimal window for stress detection. For each window size, we computed six features–mean, standard deviation, skewness, kurtosis, minimum, and maximum from each of the raw signals-RESP, ECG, EMG, EDA, and TEMP derived from the chest-worn device. We computed $5 \times 6 = 30$ features, which form the feature vector. We used these features as input to the classifiers. We used 'Baseline', 'Amusement' sessions for non-stress labels, and 'Stress' session for stress labels. The models learn the relationship between predictors and corresponding labels during the training phase, which they apply to predict stress vs. non-stress during the testing phase. The workflow diagram of our work is presented in Fig. 1.

C. *Classifiers (KNN, DT, XGBoost, LightGBM, RF, and SVM)*

To distinguish between stress and non-stress, we explored with various genres of classical ML classifiers. An advantage of machine learning models is that they perform reasonably well on small dataset common in health care settings, comparable to deep learning models. Hence, we experimented with some popularly used machine learning models such as KNN, DT, XGBoost, LightGBM, RF, and SVM to detect stress and non-stress from data collected using the chest-worn device. The ML models were trained in a supervised setting using hand engineered statistical features e.g. mean, standard deviation, skewness, kurtosis, minimum, and maximum. Once the model was learnt, test data was utilized to verify how the model would generalize. For train and test purposes, data was randomly split into a train-test ratio of 80%-20% [39]–[41]. Various performance metrics (accuracy, F1-score, and area under the curve (AUC)) were calculated using standardized techniques [42] using models' predictions and true class labels. AUC reveals the extent to which a model can distinguish between positive and negative classes. An AUC approximating to 1 indicates the model has achieved excellent separability between the classes. Conversely, an AUC close to 0 indicates the model has largely failed to learn the underlying relationship between features and corresponding classes.

K-Nearest Neighbors Algorithm: KNN as its name suggests utilizes the information of its closest neighbors, is a popular ML model which is commonly used for both regression and classification tasks. This is a simple yet useful model employed in both supervised and unsupervised approaches. A supervised learning scenario occurs when a target is known along with its features; whereas, for unsupervised learning the target labels are unknown. KNN has several hyperparameters, optimal values of which depending on the task reduce the bias and variance of a model to make it generalizable.

Decision Tree (DT): DT is a non-linear supervised algorithm which is also used for both regression and classification without requiring explicit feature normalization. This approach builds a tree-like structure by iteratively dividing the feature space at every decision point into segments depending on values of features. A major benefit of DT is that extensive pre-processing is not required for it to handle quantitative and qualitative data.

Extreme Gradient Boosting (XGBoost): We used the XGBoost classifier with a base estimator decision tree classifier [43]. The algorithm leverages a combination of regularization techniques, tree pruning, and parallel computing to enhance predictive accuracy and prevent overfitting. We conducted a grid search approach to achieve the optimal hyper-parameters with $learning\_rate$: [0.05, 0.10]; max_depth: [3, 10, 15, 50, 100]; min_child_weight : [1, 3, 5, 10, 15, 20, 50]; $gamma$: [0.1, 0.2, 0.3, 0.4]; $colsample\_bytree$: [0.3, 0.4 , 0.7]. The grid search approach showed the optimal parameters $learning\_rate$ = 0.10, max_depth = 10, min_child_weight = 5, $gamma$ = 0.3, $colsample\_bytree$ = 0.3.

Light Gradient Boosting (LightGBM): Using ensemble learning approaches for both regression and classification tasks, LightGBM is built on a gradient-boosting algorithm. This framework can handle large-scale datasets and offers the following benefits: high performance, decreased memory utilization, faster training speed, and support for parallel learning. Financial analysis, natural language processing, computer vision, and healthcare classification are just a few industries that employ it [44].

Random Forest (RF): Random Forest is a robust ensemble learning algorithm that has gained significant attention in machine learning. In this work, we also used a parameter-optimized RF classifier [45]. Optimized values of parameters $n\_estimators$, max_depth determine how the performance of a RF classifier [45]. To optimize these parameters during training, we conducted a grid search with different range of values– $n\_estimators$ with values from 50 to 500 in increments of 50; max_depth = [5, 10, 40, 50];

$min\_samples\_split$ = [2, 5, 10], to determine the maximum accuracy. We found that $n\_estimators = 250$, max_depth = 40, $min\_samples\_split$ = 2 provides us the best parameter settings for RF for this dataset that achieved the best classification accuracy.

Support vector machine (SVM): Support vector machine is another ML algorithm frequently used in various classification tasks for its robust performance. An important aspect that drives its robustness is the use of kernel function. Other tunable parameters (e.g., C, γ) also have an impact on the performance [46]. Fixing the kernel responsible for a good performance is task dependent. Hence, a grid search approach was used to determine the best kernel, C, and γ values for the classifier to achieve maximum separation between stress and non-stress classes from the training data to be able to perform well on the test data. A five-fold cross-validation [47] was used with kernels = 'RBF', where (C, γ) was fine-tuned in the range of C = [2^{-1} to 2^{10}], γ = [2^{-4} to 2^4]. SVM used the features along with class labels to learn the support vectors. The hyperplanes fixed with the largest margin (i.e., the maximum distance between the two classes) were utilized to predict test data. We selected the best model using these experiments to apply it on the unseen test data.

We separately analyzed stress detection at subject level. To do so, we split each subject's data into training (80%) and test (20%) sets. Performance metrics were calculated using models' prediction during test and true class labels. Classifiers' accuracy and other performance metrics for subject level evaluation are shown in Fig. 4 and Table II. The XGBoost classifier showed the best classification accuracy among all across states.

III. Results

We empirically evaluated the window size that can provide the most accurate classification. We conducted stress vs. non-stress detection over a fixed window block (e.g., 100 ms, 200 ms, 300 ms, 400 ms, 500 ms, 600 ms, 700 ms, 800 ms, 900 ms, and 1000 ms). We used KNN, DT, XGboost, LightGBM, RF, and SVM classifiers to investigate which window size provides the best classification accuracy.

A. Sress detection across all subjects

Stress vs. non-stress classification accuracy of six classifiers using 500 ms and 100 ms window sizes are illustrated in Fig. 2 and Fig. 3, respectively. In addition, other performance metrics are reported in Table I. The 500 ms window size data provided the highest classification accuracy among all window sizes. On the other hand, a 100 ms window yielded the lowest classification accuracy.

For 500 ms, the KNN produced a classification accuracy of 96.93%, DT 98.42%, XGBoost 99.71%, LightGBM 99.70%, RF 99.51%, and SVM 97.28%. Remarkably, XGBoost outperforms among six classifiers with the best classification accuracy of 99.71%, AUC 99.61%, precision 99.68%, recall 99.48%, and F1-score 99.57%. On the other hand, KNN yielded the lowest accuracy of 96.93%, AUC 95.21%, precision 93.95%, recall 92.93%, and F1-score 93.45%.

For 100 ms, the KNN yielded a classification accuracy of 92.10%, DT 91.47, XGBoost 94.32%, LightGBM 95.20%, RF 94.63%, and SVM 93.11%. We found that using the 100 ms window data, lightGBM showed the highest classification accuracy of 95.20%, AUC 95.13%, precision 94.31%, recall 94.83%, and F1-score 94.11% which is slightly lower than the 500 ms window results.

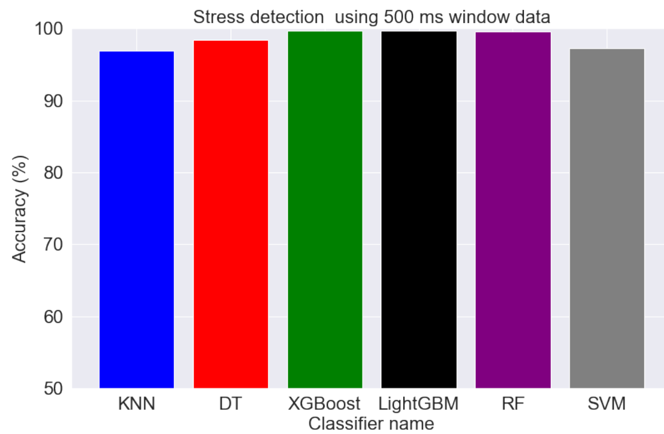

Fig. 2. Detection of stress with 500 ms time windows using KNN, DT, XGBoost, LightGBM, RF, and SVM classifiers.

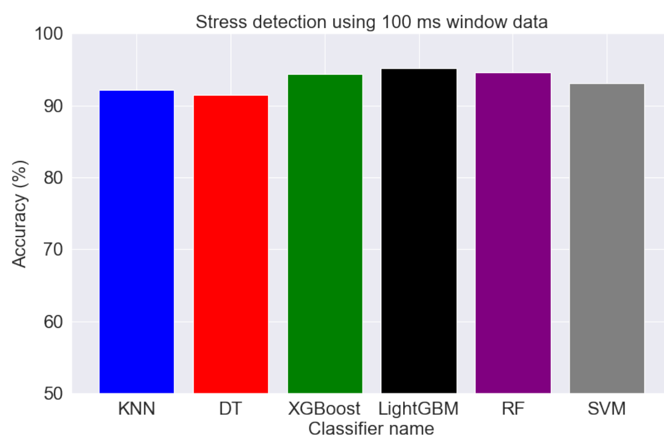

Fig. 3. Detection of stress with 100 ms time windows using KNN, DT, XGBoost, Light GBM, RF, and SVM classifiers.

B. Personalized Stress Detection

Since the 500 ms window revealed the best classification accuracy across all subjects, we used the same window size to examine subject-level evaluation. For subject-level investigation, we used the same classifiers that were used for population-level stress detection. We retrained the models and tested them with person-specific data. We report the top three classifiers' accuracy for each subject in Table II. The best classification accuracy of each subject is depicted in Fig. 4. The highest and lowest accuracies for subject-level stress detection

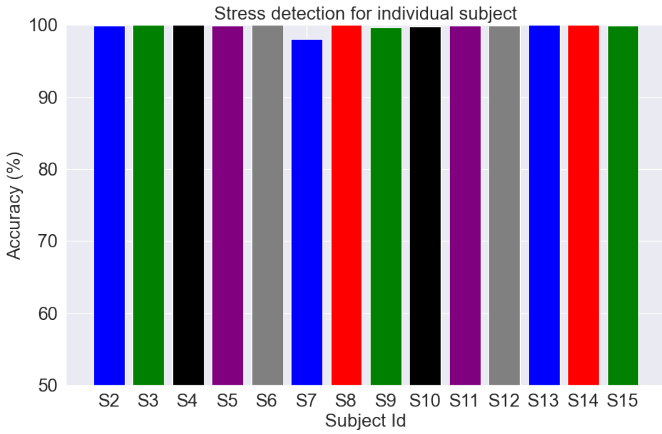

Fig. 4. Best classification accuracy with individual subject.

TABLE I.
KNN, DT, XGBOOST, LIGGHTGBM, RF, AND SVM CLASSIFIERS' PERFORMANCE METRICS (%) FOR DETECTING STRESS VS. NONSTRESS.

Classifiers name	Average measure(%)	window size stress (500 ms)	window size stress (100 ms)
KNN	Accuracy	96.93	92.10
	AUC	95.21	91.23
	Precision	93.95	90.30
	Recall	92.93	89.28
	F1-score	93.45	88.13
DT	Accuracy	98.42	91.47
	AUC	98.21	91.18
	Precision	97.31	90.32
	Recall	97.61	89.45
	F1-score	97.51	90.86
XGBoost	Accuracy	99.71	94.32
	AUC	99.61	93.25
	Precision	99.68	94.25
	Recall	99.48	98.36
	F1-score	99.57	95.25
LighGBM	Accuracy	99.70	95.20
	AUC	99.53	95.13
	Precision	99.58	94.31
	Recall	99.45	94.83
	F1-score	99.53	94.11
RF	Accuracy	99.51	94.63
	AUC	99.36	94.11
	Precision	99.69	93.85
	Recall	98.83	93.11
	F1-score	99.21	94.25
SVM	Accuracy	97.28	93.11
	AUC	96.61	92.63
	Precision	95.72	91.23
	Recall	95.71	91.11
	F1-score	95.41	91.83

of the best three models are– XGBoost (max: 99.97% and min: 97.81%), LightGBM (max: 99.95% and min: 97.85%), and RF (max: 99.92% and min: 97.82%). Our individual subject level analysis revealed that XGBoost produced the highest classification accuracy of 99.97% [AUC 99.56%, precision 99.78%, recall 99.80%, and F1-score 99.52%] and the lowest classification accuracy of 97.81% [AUC 97.77%, precision 97.13%, recall 97.50%, and F1-score 97.29%]. Hence, we examined that stress could be detected consistently well if personalized models are used unlike many other works that have proposed generalized ML models across participants with non-uniform individual performances.

TABLE II.
STRESS DETECTION ACCURACY FROM INDIVIDUAL SUBJECT LEVEL.

Subject	**XGBoost (%)**	**LightGBM(%)**	**RF(%)**
S2	99.93	99.85	99.92
S3	99.42	99.53	99.42
S4	99.97	99.80	99.92
S5	99.92	99.88	99.99
S6	99.88	99.84	99.65
S7	97.81	97.85	97.82
S8	98.07	99.68	99.45
S9	99.53	99.95	99.89
S10	99.63	99.09	99.23
S11	99.76	99.27	99.23
S13	99.89	99.29	99.68
S14	99.55	99.88	99.85
S15	99.95	99.63	99.69
S16	99.83	99.82	99.87
S17	99.77	99.88	99.89

IV. CONCLUSION

We developed a robust machine-learning framework for stress detection. Our proposed approach detected stress using a handful of statistical features with classical machine learning models. Our analysis shows stress could be detected reliably using a window size of 500 ms. Although a 500 ms window size requires fine-grained data processing, use of basic statistical features for model training can offset computational load. This potentially has positive implications for stress detection in the field specially when light-weight systems need to be deployed for rapid and accurate detection of stress. To address between-subject variability for a complex mental state like 'Stress', we developed personalized models. These personalized models yielded even better performance than population-level stress detection which may help intervention strategies catch individuals in vulnerable states more effectively. In future work, we will investigate how well the wristband sensor data can classify stress.

ACKNOWLEDGMENT

This work was supported by the Computational Data Science program and the Department of Computer Science at the Middle Tennessee State University, Murfreesboro, Tennessee, USA.

REFERENCES

[1] B. S. McEwen, "Allostasis and allostatic load: implications for neuropsychopharmacology," *Stress and the Brain*, pp. 2–18, 2013.
[2] R. M. Sapolsky, *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping*. Holt paperbacks, 2004.
[3] M. F. Dallman, "Stress-induced obesity and the emotional nervous system," *Trends in Endocrinology & Metabolism*, vol. 21, no. 3, pp. 159–165, 2010.
[4] M. R. Salleh, "Life event, stress and illness," *The Malaysian journal of medical sciences: MJMS*, vol. 15, no. 4, p. 9, 2008.
[5] J. Herbert, "Fortnightly review: Stress, the brain, and mental illness," *Bmj*, vol. 315, no. 7107, pp. 530–535, 1997.

[6] C. Drake, G. Richardson, T. Roehrs, H. Scofield, and T. Roth, "Vulnerability to stress-related sleep disturbance and hyperarousal," *Sleep*, vol. 27, no. 2, pp. 285–291, 2004.
[7] Y. H. Yau and M. N. Potenza, "Stress and eating behaviors," *Minerva endocrinologica*, vol. 38, no. 3, p. 255, 2013.
[8] K. T. Brady and S. C. Sonne, "The role of stress in alcohol use, alcoholism treatment, and relapse," *Alcohol Research & Health*, vol. 23, no. 4, p. 263, 1999.
[9] D. M. Campagne, "Stress and perceived social isolation (loneliness)," *Archives of gerontology and geriatrics*, vol. 82, pp. 192–199, 2019.
[10] S. Neupane, M. Saha, N. Ali, T. Hnat, S. A. Samiei, A. Nandugudi, D. M. Almeida, and S. Kumar, "Momentary stressor logging and reflective visualizations: Implications for stress management with wearables," *arXiv preprint arXiv:2401.16307*, 2024.
[11] J. M. Smyth and K. E. Heron, "Is providing mobile interventions" just-in-time" helpful? an experimental proof of concept study of just-in-time intervention for stress management," in *2016 IEEE Wireless Health (WH)*. IEEE, 2016, pp. 1–7.
[12] H. Gao, A. Yüce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5961–5965.
[13] S. Baltaci and D. Gokcay, "Stress detection in human–computer interaction: Fusion of pupil dilation and facial temperature features," *International Journal of Human–Computer Interaction*, vol. 32, no. 12, pp. 956–966, 2016.
[14] E. A. Sağbaş, S. Korukoglu, and S. Balli, "Stress detection via keyboard typing behaviors by using smartphone sensors and machine learning techniques," *Journal of medical systems*, vol. 44, pp. 1–12, 2020.
[15] S. Vhaduri, A. Ali, M. Sharmin, K. Hovsepian, and S. Kumar, "Estimating drivers' stress from gps traces," in *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2014, pp. 1–8.
[16] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T.-S. Chua, "Detecting stress based on social interactions in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1820–1833, 2017.
[17] M. B. Mashhadi, E. Asadi, M. Eskandari, S. Kiani, and F. Marvasti, "Heart rate tracking using wrist-type photoplethysmographic (ppg) signals during physical exercise with simultaneous accelerometry," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 227–231, 2015.
[18] R.-C. Peng, X.-L. Zhou, W.-H. Lin, Y.-T. Zhang *et al.*, "Extraction of heart rate variability from smartphone photoplethysmograms," *Computational and mathematical methods in medicine*, vol. 2015, 2015.
[19] P. Osathitporn, G. Sawadwuthikul, P. Thuwajit, K. Ueafuea, T. Mateepithaktham, N. Kunaseth, T. Choksatchawathi, P. Punyabukkana, E. Mignot, and T. Wilaiprasitporn, "Rrwavenet: A compact end-to-end multi-scale residual cnn for robust ppg respiratory rate estimation," *IEEE Internet of Things Journal*, 2023.
[20] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method," *Biomedical Signal Processing and Control*, vol. 47, pp. 196–206, 2019.
[21] C. Phillips, D. Liaqat, M. Gabel, and E. de Lara, "Wristo2: Reliable peripheral oxygen saturation readings from wrist-worn pulse oximeters," in *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2021, pp. 623–629.
[22] M. Sharma, S. Kacker, and M. Sharma, "A brief introduction and review on galvanic skin response," *Int. J. Med. Res. Prof*, vol. 2, no. 6, pp. 13–17, 2016.
[23] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: a meta-analysis and review of the literature," *Psychiatry investigation*, vol. 15, no. 3, p. 235, 2018.
[24] S. Pourmohammadi and A. Maleki, "Stress detection using ecg and emg signals: A comprehensive study," *Computer methods and programs in biomedicine*, vol. 193, p. 105482, 2020.
[25] A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, "The importance of respiratory rate monitoring: From healthcare to sport and exercise," *Sensors*, vol. 20, no. 21, p. 6396, 2020.
[26] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. McKeegan, "Skin temperature reveals the intensity of acute stress," *Physiology & behavior*, vol. 152, pp. 225–230, 2015.
[27] M. N. Hasan, S. Hamdan, S. Poudel, J. Vargas, and K. Poudel, "Prediction of length-of-stay at intensive care unit (icu) using machine learning based on mimic-iii database," in *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2023, pp. 321–323.
[28] A. Fernandes, R. Helawar, R. Lokesh, T. Tari, and A. V. Shahapurkar, "Determination of stress using blood pressure and galvanic skin response," in *2014 International Conference on Communication and Network Technologies*. IEEE, 2014, pp. 165–168.
[29] E. D. Chan, M. M. Chan, and M. M. Chan, "Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations," *Respiratory medicine*, vol. 107, no. 6, pp. 789–799, 2013.
[30] G. Singh, O. C. Phukan, and R. Kumar, "Stress recognition with multi-modal sensing using bootstrapped ensemble deep learning model," *Expert Systems*, vol. 40, no. 6, p. e13239, 2023.
[31] M. Gil-Martin, R. San-Segundo, A. Mateos, and J. Ferreiros-Lopez, "Human stress detection with wearable sensors using convolutional neural networks," *IEEE Aerospace and Electronic Systems Magazine*, vol. 37, no. 1, pp. 60–70, 2022.
[32] S. Ghosh, S. Kim, M. F. Ijaz, P. K. Singh, and M. Mahmud, "Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network," *Biosensors*, vol. 12, no. 12, p. 1153, 2022.
[33] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE Journal of Biomedical and Health Informatics*, 2023.
[34] P. Garg, J. Santhosh, A. Dengel, and S. Ishimaru, "Stress detection by machine learning and wearable sensors," in *26th International Conference on Intelligent User Interfaces-Companion*, 2021, pp. 43–45.
[35] P. Siirtola, "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, pp. 1198–1201.
[36] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.
[37] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.
[38] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade, and J. J. Gross, "Eliciting positive, negative and mixed emotional states: A film library for affective scientists," *Cognition and emotion*, vol. 30, no. 5, pp. 827–856, 2016.
[39] R. S. S. Kumari and J. P. Jose, "Seizure detection in eeg using time frequency analysis and svm," in *2011 international conference on emerging trends in electrical and computer technology*. IEEE, 2011, pp. 626–630.
[40] M. S. Mahmud, F. Ahmed, M. Yeasin, and G. M. Bidelman, "Decoding categorical speech perception from evoked brain responses," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 766–769.
[41] M. Hasan, M. Mahmud, S. Poudel, K. Donthula, and K. Poudel, "Mental workload classification from fnirs signals by leveraging machine learning," in *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2023, pp. 1–6.
[42] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.
[43] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize xgboost training parameters," *IEEE Access*, vol. 7, pp. 118 310–118 321, 2019.
[44] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (lightgbm)," *Diagnostics*, vol. 11, no. 9, p. 1714, 2021.
[45] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.
[46] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.
[47] M. Bhasin and G. Raghava, "Svm based method for predicting hla-drb1* 0401 binding peptides in an antigen sequence," *Bioinformatics*, vol. 20, no. 3, pp. 421–423, 2004.