

Personalized Stress Detection using a Lightweight Machine Learning Framework With Convenient Wrist-Worn Sensors

M. Hasan¹, S. Mithun², M. Mahmud^{3,5}, S. Baral⁴ and K. Poudel^{1,4}

1. Computational and Data Science, Middle Tennessee State University, Murfreesboro, TN, USA
2. Dept. of Computer Science, University of Memphis, TN, USA
3. Dept. of Electrical and Computer Engineering, University of Memphis, TN, USA
4. Dept. of Computer Science, Middle Tennessee State University, Murfreesboro, TN, USA
5. Div. of General Internal Medicine, Univ. of Tennessee Health Science Center, Memphis, TN, USA
mh2ay@mtmail.mtsu.edu, msaha1@memphis.edu, mmahmu10@uthsc.edu, sb2ek@mtmail.mtsu.edu, khem.poudel@mtsu.edu

Abstract— Over 70% Americans experience daily stress, making real-time, accurate stress detection crucial for timely intervention, and promoting physical and mental well-being. In this study, we detect stress using machine learning from publicly available wrist-worn sensor data. We examined both population and individual levels, focusing on simple, computation friendly statistical features for reduced processing to facilitate adoptions in real-life mHealth applications. Our framework demonstrates that stress detection achieves over 96% accuracy at both the population and individual levels, with the Extreme Gradient Boosting classifier achieving the highest predictive accuracy. We found significant variability between individuals and achieved an average accuracy of 99.63% with area under the curve (AUC) 99.23%, F1-score 99.00%, precision, and recall 99.22%. This approach offers a promising solution for real-life stress detection due to its high accuracy and simplicity.

Keywords— Stress detection, Wrist-band sensors, Machine learning, Personalization, LightGBM.

I. INTRODUCTION

Stress is a natural response, physiologically and psychologically, to demanding or threatening situations. It is primarily managed by a complex interplay of the autonomic nervous system's two branches—the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). When under stress, SNS is activated that supplies energy to the body by increasing heart rate, blood pressure, and glucose levels [1]. In contrast, the PNS aids in the body's recovery once the perceived threat subsides. Despite their distinct functions, the seamless interaction between these two systems is essential for effective bodily regulation.

While a certain amount of stress benefits cognitive function and performance [2, 3], extended and frequent exposure to stress is recognized to be harmful for both mental and physical health. Psychologically, chronic stress has been linked to anxiety, depression, and cognitive impairments [4–8]. Physiologically, it can cause cardiovascular diseases [9–13], weaken the immune system [14, 15], and lead to conditions like diabetes [16, 17] and gastrointestinal disorders [18, 19]. Moreover, the impulse to manage the discomfort of

persistent stress may lead to maladaptive coping mechanisms, which can exacerbate health problems [20]. Thus identifying the signal modalities that best capture stress signatures, while still allowing for real-time, user-friendly monitoring, is essential for accurate stress detection and effective intervention.

In the past, stress detection primarily relied on biomarkers from electrocardiogram (ECG) and respiration, which were measured with reasonable accuracy using medical-grade instruments. Prior research has consistently shown that changes in ECG derived heart rate variability (HRV) are reliable indicators of stress [21–23]. In addition, respiratory patterns, including rate and depth of breathing, alongside HRV, have been linked to stress and emotional states [24–26]. Subsequent studies on stress detection began integrating additional physiological signals to enhance detection accuracy and performance. A recent study [27] demonstrated that highly accurate personalized stress detection, with accuracy in the upper 90s, is feasible by leveraging machine learning models using multi-modal signals from chest-worn sensors. It also investigated different window sizes and found 500 ms to produce the best results. Besides, a fair amount of stress detection works exist that involve using data from non-physiological sources with promising results. For example, stress can be detected using facial expressions captured using camera [28], from contextual sources such as keyboard typing patterns [29–31], location transitions [32, 33], and logs of smartphone use [34, 35]. With the rise of social media platforms, text-based stress detection has also gained momentum [36–38]. These various methods achieve stress detection accuracy ranging from the upper 80s to mid 90s percentage. However, for real-world mHealth applications, these methods face several challenges, including but not limited to high device costs, wearability issues, privacy concerns, and intermittent monitoring. For example, collection of ECG and respiration data often requires cumbersome equipment and controlled settings, making it impractical for continuous monitoring in everyday environments. Additionally, many of these high-yield lab models developed with micro window sizes are often not scalable for real-world deployment due to their resource-intensive nature. Micro windows require

frequent processing of information, which may hamper performance and cause faster battery drainage of the host device. Hence, an optimal balance is needed between the choice of signal modalities and data processing frequency. This balance ensures ease of wearability, reliable system performance, and sustained use while still preserving effective stress detection capabilities.

Most recently, wrist-worn wearables have become quite popular for their convenience and ability to integrate seamlessly into daily life [39, 40]. These devices are equipped with sensors that measure various signals—Electrodermal Activity (EDA), also known as Galvanic Skin Response (GSR), Blood Volume Pulse (BVP), Acceleration (ACC), Heart Rate (HR), and Temperature. Their ability to collect data in real-time enables observing an individual’s physiological responses in their natural environment. Studies like [41–44] used these useful attributes to produce decent stress inferences in lab settings. However, there is still scope of improvement, specially from a modeling framework perspective. In this work, we developed lightweight ML framework to detect stress at effective granularity with reduced processing. We explored different genres of ML classifiers and feature engineering to make unobtrusive, continuous stress monitoring viable in field settings.

Our main contributions are as follows:

- Window size may affect the host device’s performance due to frequency of data processing and feature computation. We empirically identified an optimal, sufficiently large window size for best stress detection accuracy.
- We explored population-level stress detection using simple statistical features and found that Extreme Gradient Boosting (XGBoost) achieved the highest accuracy at 96.39%.
- We used Shapley values for feature explainability to identify the nine most important features out of all eighteen, which contribute most significantly to the model’s predictions. The population model’s performance improved to 97.53% when these nine features were only used.
- Due to large variability between subjects in stress reactivity, we investigated personalized stress detection. All participants registered a stress classification accuracy of 99% with one exception of 93%. The XGBoost achieved the best mean accuracy of 99.24% across subjects with a standard deviation of 1.77%.

The remainder of the paper is structured as follows: Section II covers the methodology, followed by the results in Section III. Finally, Section IV concludes the paper with a summary of the key findings.

II. MATERIALS & METHODS

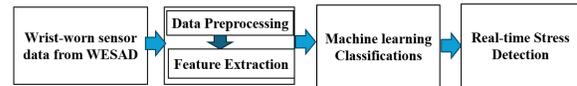


Figure 1. The process of stress detection from wrist-worn sensors.

II-A. Participants and Dataset Description

Stress detection has greatly progressed over the years due to the availability of high-quality public datasets. WESAD (Wearable Stress and Affect Detection) [45] dataset offers a comprehensive collection of physiological signals captured from both chest and wrist devices. To evaluate the proposed stress detection method, we have utilized this dataset, which includes data from 15 participants (12 males and 3 females). Participants were equipped with a RepsiBAN Professional and an Empatica E4 on chest and wrist respectively to capture stress and non-stress responses. In this study, we focused on detecting stress using only wrist-based sensors that recorded Blood Volume Pulse (BVP), Electrodermal Activity (EDA), and Skin Temperature (TEMP) at sampling rates of 64, 4, and 4 Hz, respectively, as wrist wearables are becoming more prevalent and widely adopted.

II-B. Tasks Description

To evoke diverse and targeted emotional responses, three primary sessions were conducted: Baseline, Stress, and Amusement. For the baseline session, participants spent about 20 minutes reading magazines maintaining a neutral posture by either sitting or standing at a table. The stress session, approximately 10 minutes long, was split into two equal phases to administer the Trier Social Stress Test [46]. In the first phase, participants engaged in public speaking, discussing their strengths and weaknesses before a panel. In the second phase, they performed a mental arithmetic task, counting down from 2023 to zero in increments of 17, restarting upon any error. The amusement session involved participants watching 11 humorous videos, totaling 6.5 minutes, selected from a corpus [47] and the authors’ preferences. To help participants relax, a meditation session was held following both stress and amusement sessions. Moreover, to ensure sufficient recovery from physiological response, an additional 10 minutes rest period was included right after the stress session. However, as rest is part of recovery, we use 10 minutes of stress session as stress labels and 26.5 (20+6.5) minutes of baselines and amusement sessions as non-stress labels. To obtain subjective appraisals of the conducted sessions, participants were asked to provide self-reports, to assess validity of the sessions from psycho-physiological standpoint, which added further depth to the dataset.

II-C. Feature Extraction

We experimented with different non-overlapping window sizes (e.g., 10 s, 20 s, and 30 s) 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: BVP, EDA, and TEMP derived from the wrist-worn device. Different sampling rates yielded different data points; we synchronized and considered unequal samples that fall on the same time window. We computed $3 \times 6 = 18$ features, which formed a feature vector. We used these features as input to the classifiers. We used ‘Baseline’, and ‘Amusement’ sessions for non-stress labels, and ‘Stress’ session for stress label. The aim was to ensure the window size was not too small, such as 1 second, which would require frequent feature computation, while still maintaining stress detection performance in the upper 90th percentile. This was further supported by selecting computationally efficient statistical features. 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 data were z-score normalized before submitting to the classifiers to ensure all features were on a common scale range. The process of stress detection from wrist-worn sensors is illustrated in Figure 1.

II-D. Classifiers (KNN, DT, XGBoost, LightGBM, RF, and SVM)

We explored various genres of classical ML classifiers to detect stress vs. non-stress. The advantage of ML models is that they perform reasonably well on small datasets common in healthcare settings, comparable to deep learning models. Hence, we experimented with some popularly used ML 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 trained, test data was utilized to verify how the model would generalize. This is an imbalance [stress (986) vs. non-stress (2293) ratio = 1 : 2.30] dataset. We randomly split into a train-test ratio of 80%-20% [48–50]. We reported the classification performance based on the test data that models never have seen. Various performance metrics (accuracy, F1-score, and area under the curve (AUC)) were calculated using standardized techniques [51] using models’ predictions and true class labels. We reported weighted performance measure values. AUC reveals the extent to which a model can distinguish between positive and negative classes. An AUC near 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. For an imbalance classification, precision and recall play an important role to understand the model’s robustness. If the precision and recall are closer to 1, it means an excellent model.

K-Nearest Neighbors (KNN) Algorithm:

KNN is a popular ML model commonly used for regression and classification tasks that utilize the information of its closest neighbors. This is a simplest algorithm to implement and useful 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. We fixed the KNN hyper-parameters *metric = minkowski*, *n_neighbors = 2*, *weights = distance*, *p = 2*.

Decision Tree (DT):

DT is a nonlinear supervised algorithm that can be 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 feature values. A major benefit of DT is that it does not require extensive pre-processing to handle quantitative and qualitative data. We set the hyper-parameters *criterion = entropy*, *splitter = best*, *max_depth = 50*.

Extreme Gradient Boosting (XGBoost):

We used the XGBoost classifier with a base estimator DT classifier [52]. The algorithm leverages 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, 0.20, 0.30]; *max_depth*: [5, 10, 20, 30, 50, 100, 200]; *min_child_weight* : [1, 5, 10, 15, 20, 25, 50]; *gamma* : [0.1, 0.2, 0.3, 0.4, 0.5, 0.6]; *colsample_bytree*: [0.3, 0.4, 0.5, 0.6, 0.7]. The grid search approach showed the optimal parameters *learning_rate* = 0.20, *max_depth* = 50, *min_child_weight* = 10, *gamma* = 0.2, *colsample_bytree* = 0.5.

Light Gradient Boosting Machine (LightGBM):

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 [53]. In our analysis, we set the hyper-parameters *colsample_bytree* = 0.7, *learning_rate* = 0.3, *max_depth* = 100, *min_child_weight* = 0.5, *n_estimators* = 200, *verbose* = -1.

Random Forest (RF):

RF is a robust ensemble learning algorithm that has gained significant attention in ML. In our work, we also used a parameter-optimized RF classifier [54]. Optimized values of parameters $n_estimators$, max_depth determine the performance of a RF classifier [54]. During training, we conducted a grid search to optimize the hyper-parameters. We considered hyper-parameter values- $n_estimators$ with the range from 25 to 500 in increments of 25; $max_depth = [5, 10, 20, 30, 40, 50]$; $min_samples_split = [2, 5, 10, 20]$, to determine the maximum accuracy. We found that $n_estimators = 225$, $max_depth = 10$, $min_samples_split = 5$ provides us the best parameter settings for RF for this dataset that achieved the best classification accuracy.

Support vector machine (SVM):

SVM is another ML algorithm widely used in classification and regression tasks for its robust performance. An important aspect that drives its robustness is the use of kernel tricks. Other tunable hyper-parameters (e.g., C , γ) also have an impact on the performance [55]. Fixing the kernel responsible for a good performance is task-dependent. Hence, a grid search approach was used to determine the suitable kernel, C , and γ values for the classifier to achieve maximum separation between stress and non-stress classes from the training data to perform well on the test data. A five-fold cross-validation [56] was used with kernels = ‘RBF’, where (C, γ) was fine-tuned in the range of $C = [2^{-1} \text{ to } 2^{20}]$, $\gamma = [2^{-2} \text{ to } 2^6]$. 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 ($C = 10$, $kernel = rbf$, $degree = 3$, $gamma = 0.1$) using these experiments to apply it to the unseen test data.

III. RESULTS

We conducted stress detection from wrist sensors using different window-size data at both levels (e.g., population and individual). We used parameter-optimized ML classifiers and compared the results across classifiers and window sizes.

III-A. Detection of Stress at the Population Level

First, we detected stress versus non-stress at the population level with window sizes of 10 s, 20 s, and 30 s. Classifiers’ performance for the population level stress detection is presented in Figure 2 and Table 1. The KNN classifier yielded the lowest classification accuracy across all window sizes (82.49% for 10 s, 82.03% for 20 s, and 83.48% for 30 s). XGBoost, LightGBM, and Random Forest (RF) classifiers demonstrated higher accuracy across all the windows compared to KNN, DT and SVM. Interestingly, the XGBoost classifier achieved the best performance with a 10 s window, demonstrating

an accuracy of 96.39%, AUC 95.82%, F1-score 94.58%, precision 93.45%, and recall 95.26%. However, when the window size was increased to 20 s, accuracy and F1-score dropped by 4% and 9%, respectively. A similar decline in performance was observed with the 30 s window.

This suggests that the 10 s window captures the stimuli intensity most effectively. The 10 s data maintains the optimal balance between capturing meaningful patterns and reducing noise. Our result corroborates that a smaller window like 10 s, provides finer granularity, capturing short-term fluctuations in stress levels [44, 57].

For better insights into the model’s performance, we applied Shapley Additive Explanations (SHAP) [58] over the XGBoost classifier to explain feature importance. The top-ranked 9 features (half of the total 18 features) are shown in Figure 4. Based on these top-ranked 9 features, XGBoost demonstrated an improved accuracy of 97.53%, AUC of 97.20%, F1-score of 96.23%, precision 96.19%, and recall 96.02%. Despite 50% reduction in number of features, the classification accuracy improved by 1.18%. This improvement is likely because including a larger number of features, especially less relevant ones, can lead to overfitting. As a result, the model using all the 18 features was less generalized and performed slightly worse.

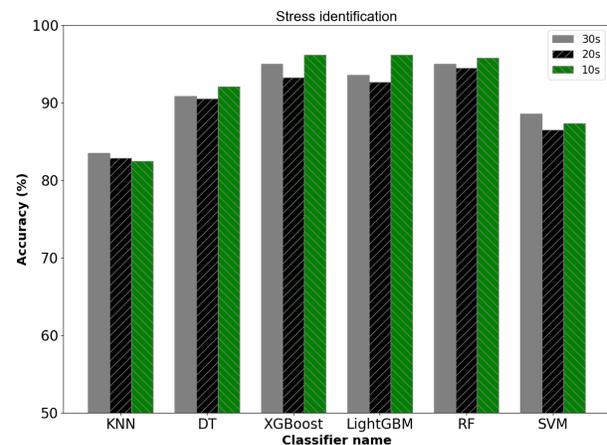


Figure 2. Stress detection with 10 s windows using KNN, DT, XGBoost, LightGBM, RF, and SVM classifiers.

III-B. Stress Detection at the individual level.

We separately conducted stress detection at the subject level. This analysis followed a similar approach to the population-level but was applied within subject-level data. Since population-level analysis identified 10 s non-overlapping window as the optimal size, we used the same window size for individual-level stress detection with the same classifiers. The best three classifiers’ accuracy is reported in Table 2. The average accuracy over the three best classifiers at subject-level is shown

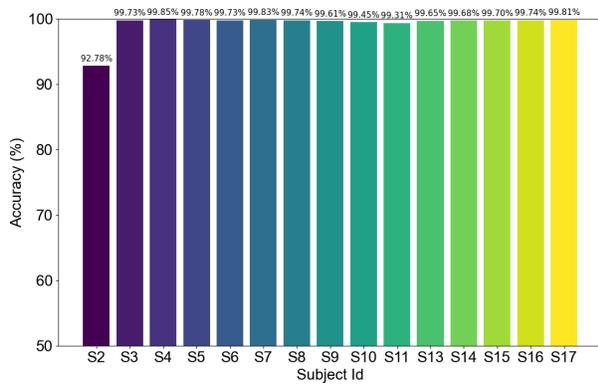


Figure 3. Highest classification accuracy with individual subject stress detection with 10 s window data. X-axis label ‘S’ indicates subject, and the number indicates data collection serial number of the subject; S1 and S12 were excluded because there were malfunctions during data collection.

Table 1. Population level stress detection performance metrics (%) for KNN, DT, XGBoost, LightGBM, RF, and SVM classifiers.

Classifiers name	Average measure(%)	window size stress (10 s)	window size stress (20 s)	window size stress (30 s)
KNN	Accuracy	83.49	82.03	83.48
	AUC	78.98	79.10	78.51
	Precision	74.48	65.21	71.92
	Recall	69.19	71.42	67.21
	F1-score	71.74	68.18	69.49
DT	Accuracy	91.15	89.82	89.88
	AUC	90.01	89.13	86.79
	Precision	85.58	80.89	85.23
	Recall	87.20	85.71	78.68
	F1-score	86.38	82.75	81.64
XGBoost	Accuracy	96.39	92.53	93.11
	AUC	95.82	91.15	90.21
	Precision	93.45	84.09	91.07
	Recall	95.26	88.09	83.60
	F1-score	94.58	86.04	87.17
LighGBM	Accuracy	96.18	92.63	93.57
	AUC	95.57	92.32	90.02
	Precision	94.28	81.91	94.33
	Recall	93.83	91.66	81.96
	F1-score	94.06	86.51	87.71
RF	Accuracy	96.03	93.86	95.87
	AUC	95.58	92.75	93.12
	Precision	93.42	86.36	98.14
	Recall	94.31	90.47	86.88
	F1-score	93.86	88.37	92.17
SVM	Accuracy	88.34	86.50	88.13
	AUC	85.06	83.91	84.01
	Precision	81.37	71.17	83.33
	Recall	78.67	78.57	73.77
	F1-score	80.00	75.00	78.26

in Figure 3. Performance was quite similar across the classifiers, with mean accuracies ranging from 99.24% to 99.16%. Stress detection for most subjects achieved 99.98% accuracy, 99.96% precision, and 99.96% recall across all classifiers. One subject, however, had a lower accuracy of 92.85%, likely due to noisier data. Still, XGboost classifier stood out with the highest mean accuracy of 99.24% and a standard deviation of 1.77%.

Table 2. Subject level stress detection accuracy for top 3 classifiers.

Subject	XGBoost (%)	LightGBM(%)	RF(%)
S2	92.85	92.70	92.80
S3	99.70	99.88	99.62
S4	99.75	99.88	99.93
S5	99.69	99.77	99.88
S6	99.77	99.88	99.55
S7	99.81	99.85	99.83
S8	99.77	99.66	99.78
S9	99.48	99.60	99.75
S10	99.73	99.19	99.43
S11	99.88	98.33	99.73
S13	99.69	99.49	99.78
S14	99.58	99.82	99.63
S15	99.51	99.89	99.70
S16	99.73	99.72	99.77
S17	99.71	99.81	99.90
Average	99.24	99.16	99.21
Std	1.77	1.83	1.79

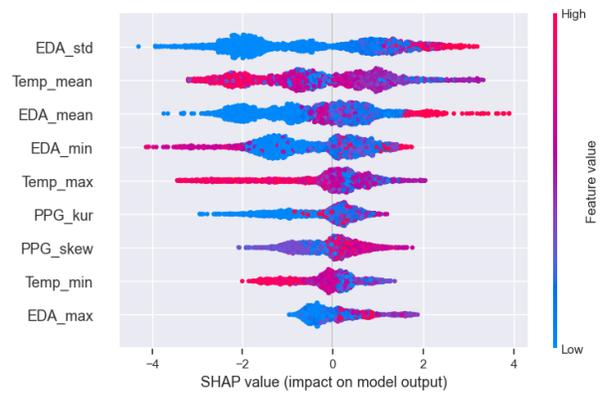


Figure 4. Density scatter plot of SHAP values obtained from XGBoost classifier. A higher value indicates the most importance feature.

We also evaluated stress detection using the top 9-ranked features identified at the population level. Most subjects achieved almost perfect accuracy (100%), with the majority showing 99.98% accuracy, AUC 99.56%, F1-score 99.48%, precision and recall 99.52%. Only three subjects had slightly lower accuracies, ranging from 96% to 98%.

IV. CONCLUSIONS

We developed an efficient ML framework for stress detection from wrist-sensor data. Our analysis shows that stress detection is more accurate based on 10 s non-overlapping windows across different classifiers. Notably, the XGBoost classifier achieved the highest classification accuracy at both the population and individual levels. Our feature selection analysis indicates that the model can predict most robustly from nine features only. However, since this analysis is based on a lab setting, performance in real-world environments may be less accurate. Our work has a limitation: we could not determine why one subject showed approximately

5% lower performance than others. Further analysis is needed to investigate this.

ACKNOWLEDGMENTS

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

REFERENCES

- [1] L. Palego, G. Giannaccini, and L. Betti, "Neuroendocrine response to psychosocial stressors, inflammation mediators and brain-periphery pathways of adaptation," *Central Nervous System Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Central Nervous System Agents)*, vol. 21, no. 1, pp. 2–19, 2021.
- [2] B. S. McEwen, "Allostasis and allostatic load: implications for neuropsychopharmacology," *Stress and the Brain*, pp. 2–18, 2013.
- [3] R. M. Sapolsky, *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping*. Holt paperbacks, 2004.
- [4] J. Herbert, "Fortnightly review: Stress, the brain, and mental illness," *Bmj*, vol. 315, no. 7107, pp. 530–535, 1997.
- [5] 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.
- [6] Y. H. Yau and M. N. Potenza, "Stress and eating behaviors," *Minerva endocrinologica*, vol. 38, no. 3, p. 255, 2013.
- [7] 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.
- [8] D. M. Campagne, "Stress and perceived social isolation (loneliness)," *Archives of gerontology and geriatrics*, vol. 82, pp. 192–199, 2019.
- [9] F. Sparrenberger, F. Cichelero, A. Ascoli, F. Fonseca, G. Weiss, O. Berwanger, S. Fuchs, L. Moreira, and F. Fuchs, "Does psychosocial stress cause hypertension? a systematic review of observational studies," *Journal of human hypertension*, vol. 23, no. 1, pp. 12–19, 2009.
- [10] M. Esler, N. Eikelis, M. Schlaich, G. Lambert, M. Alvarenga, T. Dawood, D. Kaye, D. Barton, C. Pier, L. Guo *et al.*, "Chronic mental stress is a cause of essential hypertension: presence of biological markers of stress," *Clinical and Experimental Pharmacology and Physiology*, vol. 35, no. 4, pp. 498–502, 2008.
- [11] M.-Y. Liu, N. Li, W. A. Li, and H. Khan, "Association between psychosocial stress and hypertension: a systematic review and meta-analysis," *Neurological research*, vol. 39, no. 6, pp. 573–580, 2017.
- [12] M. Kivimäki and A. Steptoe, "Effects of stress on the development and progression of cardiovascular disease," *Nature Reviews Cardiology*, vol. 15, no. 4, pp. 215–229, 2018.
- [13] I. Almadani, M. Abuhussein, and A. L. Robinson, "Yolov8-based estimation of estrus in sows through reproductive organ swelling analysis using a single camera," *Digital*, vol. 4, no. 4, pp. 898–913, 2024.
- [14] F. S. Dhabhar, "Effects of stress on immune function: the good, the bad, and the beautiful," *Immunologic research*, vol. 58, pp. 193–210, 2014.
- [15] L. Stojanovich and D. Marisavljevic, "Stress as a trigger of autoimmune disease," *Autoimmunity reviews*, vol. 7, no. 3, pp. 209–213, 2008.
- [16] R. A. Hackett and A. Steptoe, "Type 2 diabetes mellitus and psychological stress—a modifiable risk factor," *Nature Reviews Endocrinology*, vol. 13, no. 9, pp. 547–560, 2017.
- [17] F. Pouwer, N. Kupper, and M. C. Adriaanse, "Does emotional stress cause type 2 diabetes mellitus? a review from the european depression in diabetes (edid) research consortium," *Discovery medicine*, vol. 9, no. 45, pp. 112–118, 2010.
- [18] G. P. Chrousos, "Stress and disorders of the stress system," *Nature reviews endocrinology*, vol. 5, no. 7, pp. 374–381, 2009.
- [19] T. Vanuytsel, S. Van Wanrooy, H. Vanheel, C. Vanormelingen, S. Verschuere, E. Houben, S. S. Rasoel, J. Tóth, L. Holvoet, R. Farré *et al.*, "Psychological stress and corticotropin-releasing hormone increase intestinal permeability in humans by a mast cell-dependent mechanism," *Gut*, vol. 63, no. 8, pp. 1293–1299, 2014.
- [20] N. Gomes, M. Pato, A. R. Lourenco, and N. Datia, "A survey on wearable sensors for mental health monitoring," *Sensors*, vol. 23, no. 3, p. 1330, 2023.
- [21] E. T. Attar, V. Balasubramanian, E. Subasi, and M. Kaya, "Stress analysis based on simultaneous heart rate variability and eeg monitoring," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–7, 2021.
- [22] S. Huang, J. Li, P. Zhang, and W. Zhang, "Detection of mental fatigue state with wearable eeg devices," *International journal of medical informatics*, vol. 119, pp. 39–46, 2018.
- [23] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term hrv features as surrogates of short term hrv: A case study on mental stress detection in real life," *BMC medical informatics and decision making*, vol. 19, pp. 1–13, 2019.
- [24] 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.
- [25] 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.
- [26] J. Upadhya, K. Poudel, and J. Ranganathan, "A comprehensive approach to early detection of workplace stress with multi-modal analysis and explainable ai," *Proceedings of the 2024 Computers and People Research Conference*, 2024, pp. 1–9.
- [27] M. N. Hasan, M. Saha, M. S. Mahmud, S. Poudel, S. Wagle, and K. Poudel, "Personalized stress detection from chest-worn sensors by leveraging machine learning," *2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC)*. IEEE, 2024, pp. 1–6.
- [28] H. Gao, A. Yüce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5961–5965.
- [29] E. A. Sağbağ, 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.
- [30] L. Pepa, A. Sabatelli, L. Ciabattini, A. Moneriu, F. Lamberti, and L. Morra, "Stress detection in computer users from keyboard and mouse dynamics," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 1, pp. 12–19, 2020.
- [31] Y. M. Lim, A. Ayesh, and M. Stacey, "The effects of typing demand on emotional stress, mouse and keystroke behaviours," *Intelligent Systems in Science and Information 2014: Extended and Selected Results from the Science and Information Conference 2014*. Springer, 2015, pp. 209–225.
- [32] S. Vhaduri, A. Ali, M. Sharmin, K. Hovsepian, and S. Kumar, "Estimating drivers' stress from gps traces," *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2014, pp. 1–8.

- [33] D. H. Epstein, M. Tyburski, W. J. Kowalczyk, A. J. Burgess-Hull, K. A. Phillips, B. L. Curtis, and K. L. Preston, "Prediction of stress and drug craving ninety minutes in the future with passively collected gps data," *NPJ digital medicine*, vol. 3, no. 1, p. 26, 2020.
- [34] N. Yamamoto, K. Ochiai, A. Inagaki, Y. Fukazawa, M. Kimoto, K. Kiri, K. Kaminishi, J. Ota, T. Okimura, Y. Terasawa *et al.*, "Physiological stress level estimation based on smartphone logs," *2018 eleventh international conference on mobile computing and ubiquitous network (ICMU)*. IEEE, 2018, pp. 1–6.
- [35] F. Wang, Y. Wang, J. Wang, H. Xiong, J. Zhao, and D. Zhang, "Assessing mental stress based on smartphone sensing data: an empirical study," *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCCom/IOP/SCI)*. IEEE, 2019, pp. 1031–1038.
- [36] 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.
- [37] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine learning driven mental stress detection on reddit posts using natural language processing," *Human-Centric Intelligent Systems*, vol. 3, no. 2, pp. 80–91, 2023.
- [38] P. Kvtkn and T. Ramakrishnu, "A novel method for detecting psychological stress at tweet level using neighborhood tweets," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6663–6680, 2022.
- [39] A. Ometov, V. Shubina, L. Klus, J. Skibińska, S. Saafi, P. Pascacio, L. Flueratoru, D. Q. Gaibor, N. Chukhno, O. Chukhno *et al.*, "A survey on wearable technology: History, state-of-the-art and current challenges," *Computer Networks*, vol. 193, p. 108074, 2021.
- [40] N. L. Kazanskiy, S. N. Khonina, and M. A. Butt, "A review on flexible wearables-recent developments in non-invasive continuous health monitoring," *Sensors and Actuators A: Physical*, p. 114993, 2024.
- [41] 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.
- [42] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress detection using wearable physiological sensors," *Artificial Computation in Biology and Medicine: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2015, Elche, Spain, June 1-5, 2015, Proceedings, Part I 6*. Springer, 2015, pp. 526–532.
- [43] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, 2019.
- [44] M. Zubair and C. Yoon, "Multilevel mental stress detection using ultra-short pulse rate variability series," *Biomedical Signal Processing and Control*, vol. 57, p. 101736, 2020.
- [45] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.
- [46] 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.
- [47] 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.
- [48] R. S. S. Kumari and J. P. Jose, "Seizure detection in eeg using time frequency analysis and svm," *2011 international conference on emerging trends in electrical and computer technology*. IEEE, 2011, pp. 626–630.
- [49] M. S. Mahmud, F. Ahmed, M. Yeasin, and G. M. Bidelman, "Decoding categorical speech perception from evoked brain responses," *2020 IEEE Region 10 Symposium (TENSYP)*. IEEE, 2020, pp. 766–769.
- [50] M. Hasan, M. Mahmud, S. Poudel, K. Donthula, and K. Poudel, "Mental workload classification from fnirs signals by leveraging machine learning," *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2023, pp. 1–6.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.
- [55] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.
- [56] 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.
- [57] M. MK *et al.*, "Heart rate variability features for different stress classification," *Bratislava Medical Journal/Bratislavské Lekárske Listy*, vol. 121, no. 9, 2020.
- [58] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.