Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 02, Issue. 02, December 2024

RESEARCH ARTICLE OPEN ACCESS

Enhancing Stress Detection Employing Physiological Signals from the WESAD Dataset: A Machine Learning Approach with SMOTE

G Ramasubba Reddy[1]. J Jagadeswara Reddy[1] . M V Subba Reddy[1] . I Sravani[1] . V Kishen Ajay Kumar[2]

[1]Department of Computer Science & Engineering,
Sai Rajeswari Institute of Technology, Proddatur, Andhra Pradesh, India
[2]Department of Electronics and Communication Engineering,
Institute of Aeronautical Engineering, Hyderabad, Telangana, India.

DOI: **10.5281/zenodo.14566207**
Received: 13 September 2024 / Revised: 11 November 2024 / Accepted: 02 December 2024
©Milestone Research Publications, Part of CLOCKSS archiving

Abstract – Numerous medical disorders, including diabetes, heart disease, and hypertension, are significantly influenced by stress. Physiological markers for real-time stress detection are becoming more popular as wearable health monitoring devices are widely used. This research uses the WESAD (Wearable Stress and Affect Detection) dataset, which contains multimodal physiological data, including ECG, EDA, EMG, respiration, and temperature, to predict stress levels. We apply four machine learning classifiers to this dataset, focusing on addressing the class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The outcomes illustrate that Decision Trees outperform other classifiers with an accuracy of 96.27%. For future work, efforts can be directed towards incorporating additional modalities, such as EEG and eye-tracking data, to improve stress detection accuracy further. Longitudinal data collection could also help understand stress over comprehensive periods, providing insights into chronic stress patterns.

Index Terms – Stress Detection, WESAD Dataset, Machine Learning, Synthetic Minority Oversampling Technique (SMOTE), Wearable Health Monitoring, Health Monitoring Systems

MILESTONE
RESEARCH.IN
OPEN ACCESS
Electronic ISSN: **2584-072X**

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 02, Issue. 02, December 2024

I. INTRODUCTION

Stress is widespread and can be described as the body's response to challenging or demanding circumstances (Chu, 2024). Long-term or chronic stress can negatively influence people's cognitive and physical health, even though stress is a normal and adaptive response meant to assist people in overcoming risks or challenges (Vetrivel, 2024). According to the American Psychological Association (APA), stress is a normal response to the demands of daily living (Zayas, 2024). The WHO reported that 970 million people suffered from mental health problems in 2019. The Comprehensive Mental Health Action Plan 2013–2030 desires to incorporate mental health into primary care in 80% of countries by 2030 and expand service coverage by 50%. However, when it becomes extreme, it can cause anxiety, mental stress, and a host of physical symptoms that can disrupt day-to-day functioning (Ahmed, 2024). The significance of efficient stress detection and management is highlighted by the relation between chronic stress and several health problems, including heart disease, depression, and compromised immunological function (Pokhrel). Traditional approaches to stress assessment usually depend on self-report questionnaires and clinical evaluations, which have limitations (Liu, 2024). Self-reported data may be impacted by personal mood at the time of reporting, and clinical assessments usually occur infrequently, presenting only snapshots of an individual's stress levels (Babayit, 2024). Consequently, there is an increasing demand for objective, real-time stress detection techniques. Wearable technology can continuously record physiological data representing stress responses, which has brought promising developments in this field (Bolpagni, 2024). Wearable sensors that measure skin temperature, heart rate variability (HRV), and electrodermal activity (EDA) relate to the body's stress reactions (Ardecani, 2024). The autonomic nerve system controls these physiological signals, which provide unbiased information about how people respond to stressful circumstances (Schmid, 2024).

In the last few years, machine learning (ML) has evolved into a robust model for analyzing the physiological data collected by wearable devices (Xiao, 2024). ML algorithms can process enormous volumes of physiological data, extracting meaningful patterns and features that help to identify stress (Razavi, 2024). The Wearable Sensor and Affect Detection (WESAD) dataset (Benita, 2024), (Mazumdar) was developed for stress detection with physiological measurements from wearable devices. However, existing challenges include a class imbalance between stressed and unstressed data. Class imbalances must be handled to develop precise and widely applicable stress detection algorithms. To overcome these, we employ the Synthetic Minority Over-sampling Technique (SMOTE) (Elreedy, 2024), (Kaddoura, 2024). SMOTE forms synthetic samples for the lower class, balancing the dataset and assisting the ML model to learn more from the stress-labeled data. This study improves stress detection by integrating the WESAD dataset with considerable ML and SMOTE using the WESAD dataset based on physiological signals and ECG, EDA, and respiration. To deal with data imbalance, SMOTE and diverse ML classifiers are involved, which include Decision Tree (DT), XGBoost, Logistic Regression (LR), and Linear Discriminant Analysis (LDA).

II. LITERATURE REVIEW

(A machine-learning approach for stress detection using wearable sensors in free-living environments, 2024) et al. used the SWEET dataset included 240 people’s information and evaluated several classifications, to provide an ML-based methodology. RF, KNN, SVC, DT, and XGBoost were used in this analysis. The results showed that RF performed better in binary classification without SMOTE, achieving an accuracy of 98.29% and an F1-score of 97.89%. XGBoost performed exceptionally well for three-class classification using SMOTE, obtaining a comparable F1 score and an accuracy of 98.98%. (Gedam, 2024) et al. developed stress detection by integrating wearable physiological sensors with DL models including RNN and LSTM classifiers, to analyze signals like skin temperature, GSR, and ECG. RFE helped LSTM reach a maximum accuracy of 97.51%, demonstrating its resilience in stress prediction. (Geetha, 2024) et al. improved stress categorization utilizing the MLP model with sophisticated feature analysis to raise diagnostic precision. Compared to top ML algorithms like Adaboost, RF, and Gradient Boosting. Improved MLP model produced outstanding performance metrics, including 99% over a range of stress levels. (Bajpai, 2020) et al. assessed the KNN models employing the WESAD dataset, while (Di Martino, 2020) et al. presented ensemble learning with RNN models to boost stress detection accuracy. By applying the cross-validation, they evaluated the generalization ability for individual stress predictions. C.P. (Hsieh, 2019) et al. focused on feature selection according to classifier relevancy and feature correlation, validating its classification effectiveness with XGBoost. (Rashid, 2023) et al. offered the SELF-CARE method, which integrates sensor data fusion techniques to handle varied sensing conditions. Testing with wrist and chest sensors, their model achieved 86.34% and 86.19% accuracy in three-class classification and over 94% in two-class scenarios.

A comparison of six classifiers on the WESAD dataset by (Gupta, 2023) revealed the RF classifier's superior accuracy, mainly when using chest-worn sensors, which performed better than wrist sensors with 97.15% and 95.54% accuracy, respectively. Lastly, (Ghosh, 2021) et al. embarked on class imbalance through the ADASYN technique, also used a multi-class RF classifier on the WESAD dataset, performing an overall accuracy of 97.08%, representing a refined approach to stress classification utilizing both ECG and GSR signals. Benita et al. (Benita, Stress Detection Using CNN on the WESAD Dataset, 2024) presented various stress prediction approaches that influence binary and multiclass classification models, each incorporating physiological signals and different configurations to optimize model performance and utilizing a 5-second ECG signal sampled at 200Hz to reach optimal model accuracy. A novel scoring feature for stress levels was presented, varying from 0 (no stress) to 100 (high stress), adding a unique layer of granularity to stress analysis. With an accuracy of 95.04%, the model showed a high precision of 95.27% and a specificity of 99.44% in binary classifications. (Shedage, 2024) et al. underscored the role of wearable devices in stress detection, explicitly examining the effectiveness of three physiological signals, EDA, ECG, and PPG, gathered via smartwatches employing six ML models, including SVM, KNN, and a stacking ensemble approach. They presented that EDA surpassed other signals and classification methods when integrated with the stacking ensemble method, underscoring EDA's value in real-time stress detection systems. (Quadrini, 2024) et al. presented STREDWES by encoding physiological signal fragments into images using CNN and analyzing image encoding methods. Evaluations on datasets like NEURO, SWELL, and WESAD demonstrated that STREDWES effectively captured stress markers, surpassing alternative approaches.

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 02, Issue. 02, December 2024

TABLE 1. An overview of current research on stress detection

Ref	Feature Selection	Classifiers	Accuracy
(Abd Al-Alim, A machine-learning approach for stress detection using wearable sensors in free-living environments, 2024)	SMOTE	KNN, SVC, DT, RF, XGBoost	RF: 98.29%
(Gedam, 2024)	Recursive Feature Elimination (RFE)	RNN, LSTM	LSTM with RFE: 97.51%
(Geetha, 2024)	-	Multilayer Perceptron (MLP)	99%
(Bajpai, 2020)	-	KNN	90%
(Di Martino, 2020)	-	Ensemble learners and RNNs	-
(Hsieh, 2019)	RFE	XGBoost	F1-score: 92.38%
(Rashid, 2023)	-	DT, RF, AdaBoost, LDA, KNN	94.12%
(Gupta, 2023)	-	RF	chest-worn: 97.1%
(Ghosh, 2021)	ERT	RF	97.08%
(Karthick, 2022)	-	LSTM, DNN	LSTM-86% DNN-82%

III. METHODS AND MATERIALS

The proposed methodology adopts a systematic approach for detecting stress, leveraging physiological signals alongside diverse machine learning classifiers to difference between stress and non-stress states. Figure 1 graphically represents an overview of the methodology.

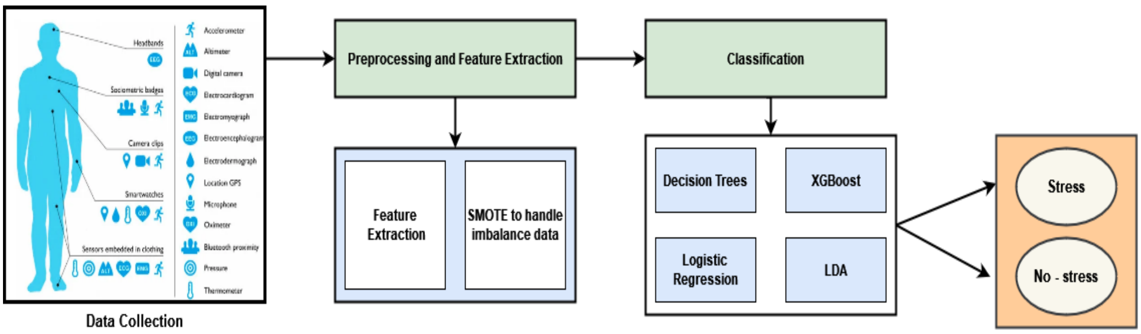

Fig. 1: Graphical representation of overall research flow

A. Dataset Description

The WESAD dataset captures data from 15 individuals, including three female participants, exposed to various emotional and stress-inducing stimuli in a controlled laboratory environment. Every participant went through three main emotional states:

- Baseline: involving a neutral reading task.
- Amusement: where they viewed humorous video clips.
- Stress: generated through the Trier Social Stress Test (TSST).

The dataset comprises physiological and motion data collected from chest-worn and wrist-worn devices. Sensor modalities include EDA, ECG, BVP, EMG, respiration (RESP), TEMP, and a 3-axis

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 02, Issue. 02, December 2024

accelerometer (ACC). For our investigation, we combine the baseline states into a single non-stress category and contrast it with the stress state in a binary classification assignment.

B. Pre-processing and Feature Extraction

Each physiological signal, excluding EMG data, was analyzed in 60-second windows for pre-processing and feature extraction. Peak detection algorithms identified individual heartbeats within raw ECG and BVP signals, enabling heart rate (HR) calculation based on the intervals between successive peaks. Mean and standard deviation were derived from these HR values. Figure 2 displays the total number of the sample in each label. A 5 Hz lowpass filter was used to process the EDA signal, which reflects sympathetic nervous system activity and heightened arousal. The EDA signal was separated into the components of tonic and phasic. SCL indicated gradual baseline shifts, and SCR captured brief reactions to stimuli. EMG signals were processed through two pathways.

First, a highpass filter removed the DC component, and the data was segmented into 5-second windows, from which statistical and frequency-domain features, including peak frequency, were extracted. Power spectral density (PSD) was also calculated across seven bands from 0 to 350 Hz. The second pathway applied a 50 Hz lowpass filter to the raw EMG signal, then segmentation into 60-second windows for extended analysis. For respiratory (RESP) signals, a bandpass filter (0.1–0.35 Hz) was used to retain only the relevant respiratory frequencies, followed by peak detection algorithms to identify minima and maxima. Min-max normalization was employed to ensure consistency across features. Given the dataset's class imbalance, the SMOTE was applied to counteract bias towards the majority class, ensuring improved model performance on imbalanced data.

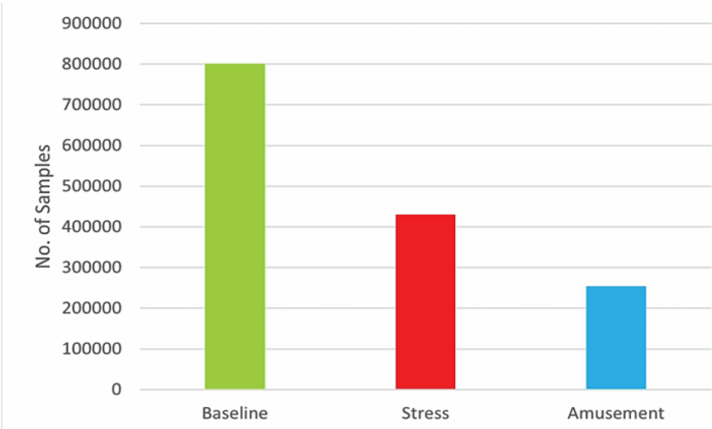

Fig. 2: Total number of Samples in each Label

C. Classification

We leveraged four ML algorithms: Decision Tree, XGBoost, Logistic Regression, and LDA to classify individuals as stressed or unstressed. Table 2 highlights each model's performance on the WESAD dataset. For the Decision Tree model, an utmost depth of 5 was set, using entropy to gauge information gain at each split. XGBoost parameters included the highest tree intensity of 5, a learning rate 0.1, and 50 estimators for stability accuracy with computational efficiency. The Logistic Regression model hired the Newton-cg solver, optimized with a 1,000-new release cap to ensure convergence. The LDA model utilized Singular Value Decomposition (SVD) because the solver successfully dealt with dimensionality

Milestone Transactions on
MEDICAL TECHNOMETRICS
Vol. 02, Issue. 02, December 2024

reduction. To provide proper evaluation, each model experienced 10-fold cross-validation to improve the reliability and generalizability of classification outcomes.

TABLE 2. Performance of the individual classifier

Classifiers	Precision	Recall	F-score	Accuracy
DT	94.76%	98.62%	96.34%	96.27%
XGBoost	85.83%	89.14%	87.16%	88.22%
Logistic Regression	90.21%	92.36%	91.19%	91.36%
LDA	70.77%	80.30%	71.71%	71.03%

IV. RESULT AND DISCUSSION

Numerous ML approaches are used to determine stress levels in the WESAD dataset, a well-liked basis for stress detection research. SMOTE was used in this study to balance the data, while classifiers such as DT, XGBoost, LR, and LDA were utilized to distinguish between stress and non-stress situations. The outcomes highlight each model's strsengths and weaknesses for essential performance metrics like F-score, recall, accuracy, and precision. DT emerged as the only classifier, accomplishing 94.76% precision, 98.62% recall, 96.34% F-1, and 96.27% accuracy. The excessive bear in mind indicates DT excels at effectively figuring out strain conditions, making it specifically helpful in situations where minimizing fake negatives is vital, including fitness tracking structures. XGBoost has proven mild overall performance, attaining 85.83% precision, 89.14% recall, 87.16% F-score, and 88.22% accuracy. While XGBoost presents robustness and performance, its lower precision than DT reveals a better possibility of false positives, which impacts its applicability in strain detection responsibilities wherein precision is a concern.LR performed well overall, delivering 90.21% precision, 92.36% recall, 91.19% F-score, and 91.36% accuracy.

The balanced performance across metrics highlights LR as a reliable choice for stress classification, particularly in resource-constrained environments where simplicity and interpretability are key. However, LDA showed the least favourable results, with 70.77% precision, 80.30% recall, 71.71% F-score, and 71.03% accuracy. Although LDA is still somewhat useful for stress detection, its much lower accuracy and precision indicate that it could be better suited for complicated datasets like WESAD, where non-linear correlations could predominate. DTs performed better than any other classifier, making them the best option for stress detection. Although LR also produced reliable and consistent results, XGBoost is a good substitute due to its processing advantages, even though it has slightly lower precision. As depicted in Figure 3, DT performed with an excellent accuracy of 96.27%, surpassing others. Additionally, using feature selection techniques could potentially improve this accuracy.

The WESAD dataset presents several chances contributions to stress detection research:

- **Increased Sample Size:** Increasing the number of participants may improve the generalizability and reliability of the data and allow for a more complete investigation of each individual's physiological differences.

- **Greater Participant Diversity:** Connecting a broad demographic scope, including varying age groups, genders, and ethnic backgrounds, should capture a unique variety of physiological responses, improving the dataset's relevance and representativeness.
- **Inclusion of Contextual Data:** Related contextual factors, like situational ramifications or environmental specifics, may enhance our view of how the outside world affects physiological states and stress reactions.
- **Longitudinal Data Collection:** Expanding the dataset to include long-term physiological data might benefit research into trends and patterns over time. This would provide insights into the dynamics of stress and affective states.
- **Expansion of Sensor Modalities:** Additional sensors, including audio, video, or eye-tracking, would improve data analysis and help for comprehensive multimodal research in affective computing and stress recognition.

With these advancements, the WESAD dataset could support more sophisticated models and increase its contributions to stress research with related applications.

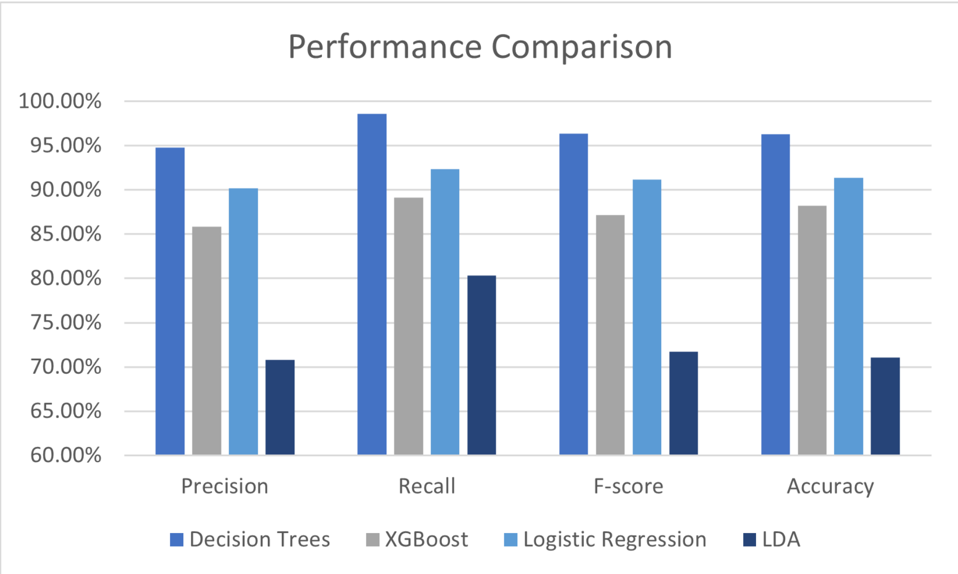

Fig. 3. Performance of the utilized classifiers

V. FUTURE WORK

Improving the WESAD dataset to cope with its current barriers could significantly enhance its value as a research resource, enabling a more profound exploration of complicated challenges in leisure reputation and affective computing. The dataset can support improving extra unique models and yield clean insights by incorporating improvements. Additionally, the WESAD framework can combine rising developments and technologies, unlocking new avenues for research.

- **Advanced Wearables Integration:** Wearable technology with existing sensors could significantly increase the range of physiological data collected by supplementing the WESAD dataset. For instance, including eye-tracking technology or sensors that measure brain activity, like EEG, would use researchers to examine cognitive functions and visual attention in addition to the physiological reactions that have already been recorded.

- **Mobile Sensing for Contextual Data:** Mobile devices like smartphones and smartwatches can integrate contextual data layers such as social communications, exercise levels, and physical position. A broader spectrum of factors impacting emotional states and behaviors should be collected to gain a more thorough understanding of emotional and physical experiences in real-world situations. Mobile sensing, which enables continuous data collection, enhances the richness of datasets for real-time analysis.
- **Multi-Modal Data Fusion:** Integrating the WESAD dataset with information from other sources, such as sentiment analysis software or social media platforms, can achieve multi-modal analysis. Researchers could combine textual or contextual data from online interactions with physiological data to better understand emotional states.This integration would boost the models' applicability in real-world scenarios, like activity detection systems, enhancing model accuracy and guaranteeing more nuanced insights.

VI. CONCLUSION

Many physiological indicators have been extensively researched in recent years to track stress levels on both a physical and mental level. Frequently, wearable sensor-based technologies use each of these signs separately. This study compares four classifiers in depth to assess how well they work in the stress detection setting, which is this study's special emphasis. The findings significantly improved when the SMOTE was used to address the problem of class imbalance and enhance model accuracy and DT outperformed the other classifiers in the test attaining remarkable accuracy.

REFERENCES

1. A machine-learning approach for stress detection using wearable sensors in free-living environments. (2024). *Computers in Biology and Medicine*, 108918.
2. Abd Al-Alim, M. a. (2024). A machine-learning approach for stress detection using wearable sensors in free-living environments. *Computers in Biology and Medicine*, 108918.
3. Ahmed, S. M. (2024). Psychological Perspectives on Work Stress: Implications for Healthcare Professionals.
4. Ardecani, F. B. (2024). Assessing the Impact of AR-Assisted Warnings on Roadway Workers' Stress Under Different Workload Conditions. *arXiv preprint arXiv:2410.14537*.
5. Babayit, O. a. (2024). The interplay of academic procrastination, self-generated stress, and self-reported bruxism among medical and dental students: a cross-sectional study. *BMC psychology*, 586.
6. Bajpai, D. a. (2020). Evaluating KNN performance on WESAD dataset. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 60--62).
7. Benita, D. S. (2024). Stress Detection Using CNN on the WESAD Dataset. In *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)* (pp. 308--313).
8. Benita, D. S. (2024). Stress Detection Using CNN on the WESAD Dataset. In *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)* (pp. 308--313).
9. Bolpagni, M. a. (2024). Personalized Stress Detection Using Biosignals from Wearables: A Scoping Review. *Sensors*, 3221.
10. Chu, B. a. (2024). Physiology, stress reaction. In *StatPearls [Internet].* StatPearls Publishing.
11. Di Martino, F. a. (2020). High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks. In *2020 IEEE symposium on computers and communications (ISCC)* (pp. 1--6).

12. Elreedy, D. a. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, 4903--4923.
13. Gedam, S. a. (2024). Advancing Mental Stress Detection in Indian Housewives: A Deep Learning Approach with Wearable Physiological Sensors and Feature Selection Methods.
14. Geetha, R. a. (2024). Human Stress Detection in and Through Sleep Patterns Using Machine Learning Algorithms. *Journal of The Institution of Engineers (India): Series B*, 1--23.
15. Ghosh, S. a. (2021). A new physiology-based objective mental stress detection technique with reduced feature set and class imbalanced dataset management. In *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)* (pp. 1--6).
16. Gupta, R. a. (2023). Multimodal wearable sensors-based stress and affective states prediction model. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 30--35).
17. Hsieh, C.-P. a.-T.-K.-Y. (2019). Feature selection framework for XGBoost based on electrodermal activity in stress detection. In *2019 IEEE International Workshop on Signal Processing Systems (SiPS)* (pp. 330--335).
18. Kaddoura, S. A. (2024). Deep convolutional neural networks with genetic algorithm-based synthetic minority over-sampling technique for improved imbalanced data classification. *Appl. Soft Comput.*, 111491.
19. Karthick, T. a. (2022). Automatic stress recognition system with deep learning using multimodal psychological data. In *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)* (pp. 122--127).
20. Liu, R. T. (2024). A systematic review and Bayesian meta-analysis of 30 years of stress generation research: Clinical, psychological, and sociodemographic risk and protective factors for prospective negative life events. *Psychological Bulletin*.
21. Mazumdar, G. a. (n.d.). ML-based stress classification and detection using multimodal affect dataset. In *Advances in Networks, Intelligence and Computing* (pp. 210--221). CRC Press.
22. Pokhrel, M. K. (n.d.). Effective Strategies for Stress Management in Teaching: Enhancing Well-being and Resilience.
23. Quadrini, M. a. (2024). Stress detection with encoding physiological signals and convolutional neural network. *Machine Learning*, 1--29.
24. Rashid, N. a. (2023). Stress detection using context-aware sensor fusion from wearable devices. *IEEE Internet of Things Journal*, 14114--14127.
25. Razavi, M. a. (2024). Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: Scoping review. *JMIR Mental Health*, e53714.
26. Schmid, R. F. (2024). Individual differences in parasympathetic nervous system reactivity in response to everyday stress are associated with momentary emotional exhaustion. *Scientific Reports*, 26662.
27. Shedage, P. S. (2024). Stress Detection Using Multimodal Physiological Signals With Machine Learning From Wearable Devices. In *2024 IEEE Symposium on Computers and Communications (ISCC)* (pp. 1--6).
28. Vetrivel, S. a. (2024). Coping Mechanisms for Stress Among Students at Universities. In *Student Well-Being in Higher Education Institutions* (pp. 157--186). IGI Global.
29. Xiao, X. a. (2024). Advances in machine learning for wearable sensors. *ACS nano*, 22734--22751.
30. Zayas, V. a. (2024). Out of sight but in mind: Experimentally activating partner representations in daily life buffers against common stressors. *Emotion*.