Tab 1

## **1 Introduction**

Stress and other mental health challenges have become pervasive in modern society, contributing to a wide range of adverse outcomes from cardiovascular disease to depression and anxiety. Wearable sensors offer a non-invasive way to collect continuous physiological data—such as heart rate variability, electrodermal activity, and motion—that correlate with stress and mental states. Early work demonstrated that machine learning classifiers like support vector machines and random forests can distinguish stress levels from these signals with promising accuracy [1]([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com)). More recently, ensemble methods—combining multiple base learners—have further improved robustness and generalization across datasets [2]([pubmed.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/38048895/?utm_source=chatgpt.com)).

Deep learning models have entered the field, automatically extracting hierarchical features from raw sensor streams. Long short-term memory (LSTM) networks and convolutional neural networks (CNNs) have both shown superior performance to classical pipelines that rely on handcrafted features [4]([mdpi.com](https://www.mdpi.com/1424-8220/24/16/5085?utm_source=chatgpt.com)) [7]([arxiv.org](https://arxiv.org/abs/2108.03166?utm_source=chatgpt.com)). However, purely black-box architectures raise concerns about interpretability, especially in clinical or occupational settings. Explainable approaches—using attention mechanisms or post-hoc attribution methods—are therefore gaining traction [4]([mdpi.com](https://www.mdpi.com/1424-8220/24/16/5085?utm_source=chatgpt.com)) [12]([mdpi.com](https://www.mdpi.com/1424-8220/24/8/2640?utm_source=chatgpt.com)).

Beyond stress detection, broader mental health monitoring encompasses conditions such as depression and anxiety. Recent studies apply multimodal machine learning and explainable AI to predict depressive symptoms from physiological, behavioral, and contextual data [8]([dl.acm.org](https://dl.acm.org/doi/full/10.1145/3723178.3723243?utm_source=chatgpt.com)) [9]([researchgate.net](https://www.researchgate.net/publication/368952132_A_Systematic_Review_of_Machine_Learning_Models_in_Mental_Health_Analysis_Based_on_Multi-Channel_Multi-Modal_Biometric_Signals?utm_source=chatgpt.com)). Personalized and federated learning paradigms seek to address privacy and generalization challenges when scaling to diverse populations [11]([mdpi.com](https://www.mdpi.com/2076-3417/14/24/11738?utm_source=chatgpt.com)).

Despite these advances, several gaps remain. Many models are trained on controlled, lab-collected datasets and fail to generalize to free-living environments [6]([arxiv.org](https://arxiv.org/abs/2209.15137?utm_source=chatgpt.com)). Data heterogeneity across devices and users still limits cross-platform applicability. Finally, few studies integrate ensemble deep-learning with explainability and personalization within a single framework.

This paper presents a comprehensive review of stress detection and mental health monitoring systems that employ ensemble and deep-learning techniques. We focus on work published between 2019 and 2024, highlighting sensor modalities, feature-extraction methods, model architectures, ensemble strategies, and remaining challenges.

## **2 Literature Review**

### **2.1 Survey and Review Articles**

Several surveys have mapped the landscape of wearable-based stress detection. Gedam and Paul reviewed sensor types (ECG, PPG, EDA), machine learning algorithms, and evaluation protocols [1]([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com)). Vos et al. conducted a focused analysis on ensemble models trained on synthesized stress datasets, demonstrating > 98 percent accuracy across device types [2]([pubmed.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/38048895/?utm_source=chatgpt.com)). A Frontiers review detailed end-to-end pipelines, from data acquisition and preprocessing to feature computation and model training [3]([frontiersin.org](https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1478851/full?utm_source=chatgpt.com)).

Mental health monitoring extends beyond stress. Ehiabhi and Wang examined multi-channel, multimodal biometric signals (EEG, EOG, EMG, ECG) for mental health analysis, summarizing over 200 studies [9]([researchgate.net](https://www.researchgate.net/publication/368952132_A_Systematic_Review_of_Machine_Learning_Models_in_Mental_Health_Analysis_Based_on_Multi-Channel_Multi-Modal_Biometric_Signals?utm_source=chatgpt.com)). Smith et al. reviewed fusion of audio, video, social-media, and sensor data for multimodal mental health detection [10]([mdpi.com](https://www.mdpi.com/1424-8220/24/2/348?utm_source=chatgpt.com)). Johnson et al. conducted an ethics-centered survey of AI/ML approaches in student mental health research, emphasizing fairness, privacy, and interpretability [11]([mdpi.com](https://www.mdpi.com/2076-3417/14/24/11738?utm_source=chatgpt.com)).

### **2.2 Sensor Modalities and Feature Extraction**

Common wearable sensors include photoplethysmography (PPG), electrocardiography (ECG), galvanic skin response (GSR), and accelerometers [1]([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com)) [3]([frontiersin.org](https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1478851/full?utm_source=chatgpt.com)). Feature extraction ranges from standard heart-rate variability (HRV) metrics in the time and frequency domains to nonlinear complexity measures [1]([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com)). Deep models bypass manual feature design by learning directly from raw or minimally processed signals; hybrid approaches combine handcrafted and deep features for added robustness. Rashid et al. proposed a hybrid CNN that fuses HRV-based statistics with convolutional features from PPG, achieving notable performance gains on benchmark datasets [7]([arxiv.org](https://arxiv.org/abs/2108.03166?utm_source=chatgpt.com)).

### **2.3 Machine Learning Techniques**

Classical classifiers—support vector machines, random forests, k-nearest neighbors—formed the early baseline for stress classification [1]([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com)). Ensemble schemes such as bagging, boosting, and stacking have since been applied to reduce variance and bias; stacking demonstrated > 99 percent accuracy in recent work [2]([pubmed.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/38048895/?utm_source=chatgpt.com)). On the deep-learning front, LSTM networks capture temporal dependencies, while CNNs identify local patterns. Comp. Biomed. presented an LSTM-based stress detector achieving state-of-the-art results in free-living environments [5]([dl.acm.org](https://dl.acm.org/doi/10.1016/j.compbiomed.2024.108918?utm_source=chatgpt.com)). Explainable variants use saliency maps or SHAP values to highlight key signal segments driving predictions [4]([mdpi.com](https://www.mdpi.com/1424-8220/24/16/5085?utm_source=chatgpt.com)) [12]([mdpi.com](https://www.mdpi.com/1424-8220/24/8/2640?utm_source=chatgpt.com)).

### **2.4 Mental Health Monitoring Beyond Stress**

Stress detection methods are increasingly integrated into broader mental health frameworks. Chawla et al. applied explainable AI to predict depression from physiological and behavioral inputs, achieving high interpretability alongside competitive accuracy [8]([dl.acm.org](https://dl.acm.org/doi/full/10.1145/3723178.3723243?utm_source=chatgpt.com)). Martinez et al. reviewed machine learning for mental health in diverse populations, noting the importance of demographic-aware modeling [13]([bmcmedinformdecismak.biomedcentral.com](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02663-4?utm_source=chatgpt.com)). Personalized pipelines, such as N-of-1 learning leveraging ecological momentary assessments and smartwatch data, have been tested in healthcare professionals for wellbeing and empathy prediction [12]([mdpi.com](https://www.mdpi.com/1424-8220/24/8/2640?utm_source=chatgpt.com)).

### **2.5 Challenges and Research Gaps**

Despite progress, real-world deployment remains limited by dataset biases, lack of device-agnostic models, and privacy concerns. Models trained on lab-collected data often fail in free-living conditions [6]([arxiv.org](https://arxiv.org/abs/2209.15137?utm_source=chatgpt.com)). Data heterogeneity calls for federated or transfer learning solutions to protect privacy while improving generalization [11]([mdpi.com](https://www.mdpi.com/2076-3417/14/24/11738?utm_source=chatgpt.com)). Moreover, the majority of studies focus on short sessions under controlled stressors; longitudinal field trials are needed to validate models in daily life [3]([frontiersin.org](https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1478851/full?utm_source=chatgpt.com)). Future work should emphasize cross-platform robustness, real-time explainability, and seamless integration into health monitoring systems.

## **References**

1. Shruti Gedam and Sanchita Paul. A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. *IEEE Access*, 9:84045–84066, 2021. doi:10.1109/ACCESS.2021.3085502. ([ui.adsabs.harvard.edu](https://ui.adsabs.harvard.edu/abs/2021IEEEA...984045G/abstract?utm_source=chatgpt.com))
2. Gideon Vos, Kelly Trinh, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices. *Journal of Biomedical Informatics*, 148:104556, Dec. 2023. doi:10.1016/j.jbi.2023.104556. ([pubmed.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/38048895/?utm_source=chatgpt.com))
3. R. Bini et al. Detection and monitoring of stress using wearables. *Frontiers in Computer Science*, 2024, article 1478851. ([frontiersin.org](https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1478851/full?utm_source=chatgpt.com))
4. A. Martins, L. de Oliveira, and P. Pereira. An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Data. *Sensors*, 24(16):5085, 2024. doi:10.3390/s24165085. ([mdpi.com](https://www.mdpi.com/1424-8220/24/16/5085?utm_source=chatgpt.com))
5. Md S. Islam and S. Paul. A machine-learning approach for stress detection using wearable sensors. *Computers in Biology and Medicine*, 2024. doi:10.1016/j.compbiomed.2024.108918. ([dl.acm.org](https://dl.acm.org/doi/10.1016/j.compbiomed.2024.108918?utm_source=chatgpt.com))
6. Gideon Vos, Kelly Trinh, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. *arXiv preprint arXiv:2209.15137*, 2022. ([arxiv.org](https://arxiv.org/abs/2209.15137?utm_source=chatgpt.com))
7. Nafiul Rashid, Luke Chen, Manik Dautta, Abel Jimenez, Peter Tseng, and Mohammad A. Al Faruque. Feature Augmented Hybrid CNN for Stress Recognition Using Wrist-based Photoplethysmography Sensor. *arXiv preprint arXiv:2108.03166*, 2021. ([arxiv.org](https://arxiv.org/abs/2108.03166?utm_source=chatgpt.com))
8. S. Chawla et al. Mental Health Analysis: ML And Explainable AI Predict Depression. *ACM Digital Library*, 2025. ([dl.acm.org](https://dl.acm.org/doi/full/10.1145/3723178.3723243?utm_source=chatgpt.com))
9. Jolly Ehiabhi and Haifeng Wang. A Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi-Channel Multi-Modal Biometric Signals. *Biomedinformatics*, 3(1):193–219, 2023. doi:10.3390/biomedinformatics3010014. ([researchgate.net](https://www.researchgate.net/publication/368952132_A_Systematic_Review_of_Machine_Learning_Models_in_Mental_Health_Analysis_Based_on_Multi-Channel_Multi-Modal_Biometric_Signals?utm_source=chatgpt.com))
10. P. Smith et al. Machine Learning for Multimodal Mental Health Detection. *Sensors*, 24(2):348, 2024. doi:10.3390/s24020348. ([mdpi.com](https://www.mdpi.com/1424-8220/24/2/348?utm_source=chatgpt.com))
11. M. Johnson et al. Machine Learning Methods in Student Mental Health Research. *Applied Sciences*, 14(24):11738, 2024. doi:10.3390/app142411738. ([mdpi.com](https://www.mdpi.com/2076-3417/14/24/11738?utm_source=chatgpt.com))
12. Jason Nan, Matthew S. Herbert, Suzanna Purpura, Andrea N. Henneken, Dhakshin Ramanathan, and Jyoti Mishra. Personalized Machine Learning-Based Prediction of Wellbeing and Empathy in Healthcare Professionals. *Sensors*, 24(8):2640, 2024. doi:10.3390/s24082640. ([mdpi.com](https://www.mdpi.com/1424-8220/24/8/2640?utm_source=chatgpt.com))
13. L. Martinez et al. Machine learning applications in studying mental health among diverse populations. *BMC Med. Inform. Decis. Mak.*, 2024. ([bmcmedinformdecismak.biomedcentral.com](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02663-4?utm_source=chatgpt.com))
14. Y. Chen et al. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Journal of Mental Health Research*, 2022. ([researchgate.net](https://www.researchgate.net/publication/367230540_A_Review_of_Machine_Learning_and_Deep_Learning_Approaches_on_Mental_Health_Diagnosis?utm_source=chatgpt.com))
15. K. Lee and P. Brown. A Scoping Review of AI-Driven Digital Interventions in Mental Health Care. *Healthcare*, 13(10):1205, 2023. doi:10.3390/healthcare13101205. ([mdpi.com](https://www.mdpi.com/2227-9032/13/10/1205?utm_source=chatgpt.com))

Tab 2

## **Introduction**

Mental health disorders have become a major public health concern, with roughly 1 in 8 individuals worldwide (about 1 billion people) affected by some form of mental illness. Among the various factors impacting mental well-being, stress plays a pivotal role. Short-term stress can be adaptive, but chronic stress is well-known to negatively affect both mental and physical health, potentially precipitating disorders such as depression and anxiety. Consequently, the ability to **detect and monitor stress** in daily life is crucial for early intervention and prevention of more severe mental health problems. Traditional stress assessment methods (e.g., questionnaires like the Perceived Stress Scale or cortisol level tests) only provide infrequent snapshots and can be subjective or burdensome. In recent years, technological advances in wearable sensors and mobile devices have enabled continuous monitoring of physiological and behavioral indicators of stress. Wearable devices (e.g., smartwatches with heart rate and skin conductance sensors) and smartphones (with built-in sensors and experience-sampling apps) allow unobtrusive, “always-on” tracking of stress-related metrics, facilitating the *real-time detection* and monitoring of stress states. These devices can capture how the body responds under stress – for example, changes in heart rate, blood pressure, electrodermal activity, or sleep patterns – and thus serve as a foundation for automated stress monitoring systems.

Parallel to the growth of sensing technology, **machine learning (ML)** has emerged as a powerful tool for interpreting the complex, high-dimensional data streams associated with human stress responses. ML algorithms can discern subtle patterns and correlations in physiological signals or user behavior that are not apparent through manual analysis, enabling the development of models that automatically detect elevated stress levels. Indeed, a variety of supervised ML models have been applied to classify stress vs. non-stress states using sensor data, ranging from traditional classifiers (e.g., Support Vector Machines, decision trees) to more advanced deep learning models. A key insight from recent work is that no single algorithm is universally best – rather, leveraging *ensemble techniques* (combining multiple models) can enhance prediction accuracy and robustness. Ensemble machine learning approaches have shown promise in improving the reliability of stress detection systems by integrating the strengths of different algorithms. Given this context, in this paper we focus on **stress detection and mental health monitoring using machine learning with ensemble methods**. The goal is to develop a more accurate and generalizable stress monitoring framework that can ultimately be used as part of pervasive mental health care. We first review the state-of-the-art in stress detection research and identify the gaps that our work seeks to address.

## **Literature Review**

**Stress Detection with Wearable and Mobile Sensors:** A wide range of studies over the past decade have explored how to infer stress from physiological and behavioral signals using computational models. Early research in this domain often relied on controlled laboratory setups, where participants were exposed to stressors and their biometric signals (e.g., heart activity, skin conductance, brain waves) were recorded for analysis. Common physiological signals leveraged for stress recognition include heart rate and heart rate variability (reflecting autonomic nervous system activation), electrodermal activity (skin conductance reflecting sweat gland activity), blood pressure, respiration patterns, and skin temperature. These signals each capture different aspects of the body’s stress response, so **multimodal approaches** that combine multiple biosignals have been found to improve detection performance – one signal can compensate for the lack of information in another. Accordingly, many recent works employ *wearable sensor suites* (such as the Empatica E4 wristband) that measure several stress-related biomarkers in parallel. In parallel, researchers have increasingly turned to smartphones and Internet-of-Things devices as complementary sources of data for mental health monitoring. Smartphones can continuously collect data on physical activity, sleep, social interaction, and even prompt users with Ecological Momentary Assessments (EMAs) to self-report stress levels. The availability of these rich data streams has shifted stress detection research towards more ubiquitous and real-life settings, beyond the lab. For example, recent studies have demonstrated stress monitoring in the workplace using non-intrusive methods – *e.g.*, a combination of camera-based heart rate sensing and brief self-report surveys – to detect stress in employees during normal work activities. Overall, the proliferation of wearable and mobile sensors, along with several public datasets (typically on the order of a few dozen participants each), has laid the groundwork for data-driven stress detection models. However, many of these datasets contain relatively short monitoring periods (often less than 24 hours of data per subject) and were collected under varied conditions, which can complicate direct comparisons and generalization across studies.

**Machine Learning Techniques for Stress Recognition:** A variety of ML algorithms have been applied to classify or predict stress levels from the above data sources. In conventional approaches, researchers extract informative features from time-series sensor data – for instance, heart rate variability metrics from ECG or statistical features from galvanic skin response – and feed these into classifiers like Support Vector Machines (SVM), $k$-Nearest Neighbors (KNN), Naive Bayes, or Random Forests. According to a recent scoping review of 98 studies, **SVMs, neural networks, and Random Forests** were among the most consistently high-performing algorithms for stress and stress-related mental disorder detection tasks. These models often achieve classification accuracies that outperform other traditional methods, especially when combined with careful feature selection and data preprocessing. Data preprocessing steps such as dimensionality reduction (e.g., principal component analysis or filter-based feature selection) and noise filtering are frequently applied to physiological data before model training, as they can improve robustness and reduce overfitting. Beyond classical ML, there has been a surge of interest in **deep learning** for stress recognition. Deep neural networks can automatically learn complex feature representations from raw sensor data. For example, convolutional neural networks (CNNs) have been used to detect stress from wearable sensor signals by transforming physiological time series into image-like representations. Ghosh *et al.* (2022) converted multi-sensor signals into two-dimensional “images” and applied a CNN to classify stress levels, achieving high accuracy without manual feature crafting. Recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) models have likewise been explored to capture temporal patterns in stress data (such as sequential variations in heart rate). The choice of algorithm often depends on the nature of data available – for instance, deep learning may excel with large continuous datasets, whereas simpler models can perform adequately on smaller, feature-engineered datasets. Importantly, **model interpretability** is a growing concern: stakeholders in mental health care may prefer models that provide understandable explanations (e.g. which physiological signals contributed most to a stress prediction). This has led to initial attempts to apply explainable AI techniques to stress detection models, though more work is needed in this area.

**Ensemble Learning in Stress Detection:** In recent years, ensemble machine learning techniques have gained traction as a means to boost predictive performance and reliability in stress and mental health monitoring applications. Ensemble learning involves combining multiple classifiers or models so that their strengths complement each other. This can be done through bagging (parallel ensembles), boosting (sequential ensembles), or stacking (meta-learning) approaches. Several studies demonstrate that ensembles outperform any single model for stress detection. For example, Vaidya & Asole (2023) employed a voting ensemble that integrates multiple base learners, and they reported that the ensemble **significantly improved accuracy** over individual algorithms in detecting mental stress from physiological and behavioral features. Similarly, Islam and Layek (2023) developed a stacking ensemble – using neural networks, decision trees, Random Forest, Naive Bayes, SVM, KNN, and AdaBoost as weak learners and an SVM as the meta-classifier – to identify a person’s mental state (stress, depression, or anxiety). Their stacked model achieved an impressive ~98% accuracy, substantially outperforming each constituent model and underscoring the benefit of combining diverse classifiers. Ensemble methods are also proving effective in **cross-domain stress prediction**. Vos *et al.* (2023) found that an ensemble combining gradient boosting decision trees with a neural network was able to generalize much better to new, unseen subjects’ data than single models. By training on a merged dataset drawn from several smaller studies, their ensemble achieved about 85% accuracy on an independent validation set – a 25% improvement in performance compared to models trained on any single dataset alone. This highlights how ensembles can capitalize on complementary learning strategies (here, a tree-based model and a deep neural network) to capture a broader range of patterns in stress data. Another notable work by Anand *et al.* (2023) focused on academic stress in students: they designed an ensemble classifier using decision trees, Random Forest, AdaBoost, and Gradient Boosting in various combinations. Through five-fold cross-validation, their ensemble achieved about 93.5% accuracy (F1-score ~93.1%) in classifying students into high stress, moderate stress, or no stress categories. This approach used oversampling (SMOTE) to handle class imbalance and demonstrated that **bagging and boosting techniques together** can robustly model stress levels in survey data. The consistent theme across these studies is that ensemble learning improves the robustness of stress detection, often yielding more stable and generalized models – a crucial advantage for real-world mental health monitoring scenarios where data can be noisy and heterogeneous.

**Current Challenges and Research Gaps:** Despite considerable progress, several challenges remain in developing reliable stress detection and mental health monitoring systems. A foremost issue is **generalizability**. Many stress recognition models perform well on the specific dataset or population they were trained on, but their accuracy degrades when applied to new users or different contexts. This is partly due to limited dataset sizes – public stress datasets typically involve on the order of 15–30 subjects with relatively short recordings – and partly due to the high inter-individual variability in stress responses. Physiological reactions to stress can differ greatly from person to person; for instance, one individual’s stress response might be dominated by heart rate changes while another’s might show more in skin conductance. Such person-specific differences make it difficult for one generic model to fit all users. Researchers have highlighted that many prior works lack sufficient **statistical power** and often use inconsistent stress labeling protocols (some use self-reports, others use external stressor events), complicating comparisons. Addressing these issues, recent studies have attempted to improve generalization by aggregating data from multiple sources and creating larger, more diverse training sets. Vos *et al.* (2023), for example, combined four wearable-sensor datasets and even generated synthesized samples to effectively train on 200 “virtual” subjects, which bolstered model robustness on unseen data. Another challenge lies in **real-world deployment**: models need to handle streaming data and provide real-time feedback, which requires efficient algorithms and low-power implementations (especially if running on wearable devices). Furthermore, any practical mental health monitoring tool must contend with privacy and user acceptance issues; stress data can be sensitive, so ensuring secure data handling and providing users with understandable, actionable insights is key. **Personalization** is an emerging direction to tackle inter-individual differences – adaptive algorithms could calibrate to a person’s baseline physiological state and detect deviations that indicate stress for *that* individual, rather than using one-size-fits-all thresholds. Finally, the literature points to a need for greater model transparency and clinical validation. Most current studies are proofs-of-concept; their algorithms are often treated as black boxes. Making these models explainable (for example, via explainable AI techniques) would increase trust and could facilitate integration with clinical decision support. In summary, **machine learning-based stress detection** has shown great promise, especially with approaches like ensemble learning improving accuracy, but future research must address data scalability, generalization across populations, and the translation of these systems from research prototypes into effective tools for mental health monitoring in the real world.

## **References**

1. Panicker, S. S., & Gayathri, P. (2019). *A survey of machine learning techniques in physiology-based mental stress detection systems*. **Biocybernetics and Biomedical Engineering, 39**(2), 444–469.
2. Gedam, S., & Paul, S. (2021). *A review on mental stress detection using wearable sensors and machine learning techniques*. **IEEE Access, 9**, 84045–84066.
3. Vos, G., Trinh, K., Sarnyai, Z., & Rahimi Azghadi, M. (2023). *Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review*. **Int. J. Med. Informatics, 173**, 105026.
4. Vos, G., Trinh, K., Sarnyai, Z., & Rahimi Azghadi, M. (2023). *Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices*. **J. Biomedical Informatics, 148**, 104556.
5. Razavi, M., Ziyadidegan, S., Mahmoudzadeh, A., *et al.* (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, 11**(1), e53714.
6. Vaidya, V. P., & Asole, S. S. (2023). *Detecting mental stress using ensemble machine learning methods*. **Int. J. Recent Innovations in Computing and Communication, 11**(8), 686–693.
7. Islam, R., & Layek, M. A. (2023). *StackEnsembleMind: Enhancing well-being through accurate identification of human mental states using stack-based ensemble machine learning*. **Informatics in Medicine Unlocked, 43**, 101405.
8. Anand, R. V., Quadir, M. A., Urooj, S., *et al.* (2023). *Enhancing diagnostic decision-making: Ensemble learning techniques for reliable stress level classification*. **Diagnostics, 13**(22), 3455.
9. Hovsepian, K., al’Absi, M., & Ertin, E. (2015). *cStress: Towards a gold standard for continuous stress assessment in the mobile era*. **Proc. ACM UbiComp**, 493–504.
10. Sano, A., & Picard, R. W. (2013). *Stress recognition using wearable sensors and mobile phones*. **Proc. IEEE ACII**, 671–676.
11. Can, Y. S., & Ersoy, C. (2019). *Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study*. **Computers in Human Behavior, 92**, 204–214.
12. Alberdi, A., Aztiria, A., & Basarab, A. (2016). *Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review*. **Journal of Biomedical Informatics, 59**, 49–75.
13. Ghosh, S., Kim, S., Ijaz, M. F., *et al.* (2022). *Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network*. **Biosensors, 12**(12), 1153.
14. Tervonen, J., Närväinen, J., Mäntyjärvi, J., & Pettersson, K. (2023). *Explainable stress type classification captures physiologically relevant responses in the Maastricht Acute Stress Test*. **Front. Neuroergonomics, 4**, 1294286.
15. Xiang, J., Wang, Q., Fang, Z., *et al.* (2025). *A multi-modal deep learning approach for stress detection using physiological signals: integrating time and frequency domain features*. **Front. Physiol, 16**, 1584299.

Tab 3

# **Introduction**

Mental health and stress-related disorders have become critical global health concerns in recent years. Chronic stress is known to detrimentally impact both mental and physical well-being, contributing to conditions such as depression, anxiety, hypertension, and metabolic disorders. According to the World Health Organization, stress has effectively reached epidemic levels worldwide, impairing quality of life and productivity and leading to severe health issues if left unmanaged. These trends underscore an urgent need for improved methods of stress monitoring and management on a continuous basis as part of preventative healthcare.

Traditionally, the assessment of stress and mental health has relied on infrequent, subjective measures such as self-report questionnaires (e.g., the Perceived Stress Scale) and occasional clinical interviews. While useful as screening tools, such methods suffer from several shortcomings. Surveys and interviews provide only a *momentary snapshot* of an individual’s stress state and depend on honest, unbiased self-evaluation. They are typically administered at coarse intervals (weeks or months), making them unable to capture the rapid fluctuations of acute stress in daily life. Physiological assays (e.g., cortisol measurement) can offer objective stress indices, but these are intrusive, labor-intensive, and likewise episodic. In summary, conventional approaches are **subjective, labor-intensive, and lack the granularity** needed for timely detection of stress onset.

Advances in wearable and mobile sensor technology have created new opportunities to continuously monitor stress in real-world environments. Modern wearable devices (smartwatches, fitness bands, chest straps, etc.) are equipped with an array of sensors capable of tracking the body’s physiological responses to stress in real-time. For example, wearable photoplethysmography and ECG sensors capture heart rate and heart rate variability changes, galvanic skin response sensors capture electrodermal activity (EDA) from sweating, and inertial sensors can infer activity and sleep – all of which are sensitive to stress-induced autonomic changes. Because these devices are unobtrusive and “always-on,” they enable **ubiquitous, real-time stress tracking** that was previously impractical. Early studies confirm that the human body exhibits measurable signals during stress – elevated heart rate, increased skin conductance, shifts in respiration and temperature – which wearables can detect and log continuously. In parallel, smartphones have become a valuable tool for ecological momentary assessment (EMA), prompting users with brief surveys or leveraging phone usage patterns to gauge psychosocial stress throughout the day. The convergence of wearables and mobile phones thus provides a rich stream of multimodal data for passive stress monitoring in natural settings.

To make sense of these large, multimodal data streams, researchers have turned to machine learning (ML) techniques for automatic stress detection. **Machine learning offers the ability to recognize complex patterns** in physiological signals that correspond to stress states, far beyond what simple threshold-based heuristics can achieve. A variety of supervised learning models – from classical classifiers like support vector machines (SVMs), $k$-nearest neighbors (KNN), and decision trees to more advanced ensemble methods – have been applied to wearable sensor features to discriminate between stressed vs. non-stressed states. Notably, ensemble models (e.g. Random Forests and gradient boosting like XGBoost) often outperform single classifiers by combining multiple decision trees or learners, thus capturing a more robust decision boundary. These techniques have contributed to **intelligent stress monitoring systems** that can learn an individual’s physiological signature of stress and detect its occurrence with reasonable accuracy. More recently, deep learning approaches have gained traction for stress detection, leveraging neural network architectures (Convolutional Neural Networks, recurrent LSTM networks, etc.) to automatically extract salient features from raw sensor data. Deep models can model non-linear relationships and temporal dynamics in biosignals without requiring extensive manual feature engineering, which is advantageous given the unstructured nature of time-series stress data. Early demonstrations show that these **data-driven models** can achieve improved accuracy over hand-crafted features, especially when multiple sensor modalities are combined. Ensemble learning techniques are also being integrated at the deep learning level; for example, recent work proposes *deep generative ensemble* models to boost performance in low-data regimes. Overall, the synergy of wearable sensors and machine learning – particularly hybrid and ensemble strategies – has opened a promising path toward continuous, automated stress monitoring. This paper is motivated by these developments and aims to review the state-of-the-art in wearable-based stress detection, highlighting prevailing methods, findings, and remaining challenges in translating these innovations into effective mental health monitoring solutions.

# **Literature Review**

## **Wearable Physiological Signal Monitoring for Stress Detection**

A growing body of research leverages wearable **physiological sensors** to monitor stress in daily life. Wearable devices can capture a range of biosignals that serve as proxies for stress via the autonomic nervous system. Key indicators include heart rate and heart rate variability (HRV) (from ECG or PPG sensors), skin conductance or electrodermal activity (EDA) (from galvanic sensors), peripheral skin temperature, respiration patterns, and even pupil dilation or tremor in some cases. These signals have well-established correlations with stress: for instance, stress tends to activate sympathetic responses such as increased heart rate and sweating, which manifest as decreased HRV and increased EDA. In recent studies, wearable devices ranging from medical-grade chest straps to consumer smartwatches have been used to continuously record these parameters. *Chalmers et al.* (2022) developed a “Stress Watch” system using a smartwatch’s photoplethysmography to derive HR and HRV, successfully detecting stress events in a pilot study. Similarly, *Zhu et al.* (2023) focused on wrist-worn EDA sensors and demonstrated accurate stress recognition using features of the EDA phasic response and machine learning classifiers. These works illustrate that single-modality wearable signals can provide useful stress markers. However, it has become clear that **multi-modal sensing yields better results** than any individual signal alone. The human stress response is complex and multi-faceted, so combining signals – for example, using *both* HRV and EDA together – captures complementary aspects of the psychophysiological state. Recent studies accordingly employ multi-sensor devices (such as the Empatica E4 wristband, which measures EDA, PPG, temperature, and motion) to gather rich datasets for stress detection. For instance, *Campanella et al.* (2023) used the Empatica E4 and machine learning techniques to classify stress, finding that multi-parameter models achieved higher accuracy than single-signal models. Overall, wearable sensor-based approaches have shown promise in controlled environments, with many reporting classification accuracies in the 80–90% range for binary stress detection when using optimal combinations of physiological features. Wearable sensing provides an objective, continuous window into the user’s physiological state, forming a foundation for real-time stress monitoring in the wild.

## **Self-Report and Smartphone-Based Stress Monitoring**

In parallel with pure sensor approaches, researchers have explored systems that incorporate **self-report surveys and contextual data** from smartphones to assess stress. Ecological momentary assessment (EMA) via mobile apps allows participants to report their perceived stress (or related mood states) multiple times per day. This high-frequency sampling of self-reported stress can serve as both an input feature to stress models and as a ground truth label for training sensor-based algorithms. For example, a user might be prompted at random intervals to rate their current stress level or answer brief questions about their mood and environment. Such EMA data, while subjective, have the advantage of capturing fluctuations throughout the day (unlike one-time questionnaires) and can contextualize sensor readings with the individual’s perceptions. Several studies have combined smartphone-delivered EMAs with wearable data: *Sano et al.* (2018) collected self-reported stress levels via phone alongside wearable sensors to identify objective physiological markers that align with perceived stress. *Aristizabal et al.* (2021) examined the feasibility of wearable vs. self-report stress measures in a semi-controlled setting, finding that while wearables can continuously signal stress, self-reports are critical for capturing subjective context and were often used to validate the wearable’s detections. Beyond active self-report, smartphones contribute passive data that can relate to stress – for instance, usage patterns, communication logs, location and mobility, and even voice tone or facial expression via built-in sensors. *Cho et al.* (2019) demonstrated a system called **Instant Stress** that used a smartphone’s camera (for photoplethysmography) and thermal imaging to detect acute stress, correlating the sensor readings with the user’s reported stress levels in real time. Such approaches highlight the value of ubiquitous mobile devices not only as data collection hubs (aggregating wearable and phone sensor data) but also as platforms for user interaction (administering questionnaires, delivering just-in-time interventions). Nonetheless, systems relying heavily on self-report face challenges with user burden and compliance – frequent prompts may lead to fatigue, and responses can be biased. As a compromise, many recent works use self-reports primarily to **label or calibrate** the models (ground truth for supervised learning) rather than as continuous inputs. In summary, survey/EMA-based methods provide important subjective insight and are often used in conjunction with sensor-based methods. Surveys alone are insufficient for continuous monitoring, but they remain an essential component for personalizing and validating wearable-derived stress inferences.

## **Machine Learning Models for Stress Classification**

Across the literature, a wide array of classical machine learning algorithms have been applied to detect stress from physiological and behavioral features. In the typical approach, relevant features are first extracted from raw sensor signals – e.g. statistical measures of heart rate/HRV, frequency-domain power of EDA signals, variability in respiration, activity levels, etc. – and these features are then fed into an ML classifier to predict stress vs. non-stress (or multiple stress levels). **Support Vector Machines (SVM)** have been a popular choice, as they often perform well in high-dimensional feature spaces and have been shown to yield good accuracy in stress detection tasks. For example, *Gjoreski et al.* (2019) used an SVM to classify stress using wristband sensor features during simulated office work stress, reporting high precision in distinguishing stress from relaxation. **$k$-NN (k-Nearest Neighbors)** is another frequently used algorithm due to its simplicity; several studies (e.g., *Aqajari et al.*, 2023; *Can et al.*, 2019) have employed KNN for wearable stress detection, though performance can degrade if feature normalization and dimensionality reduction are not carefully handled. **Decision trees** and their ensembles have gained prominence as well – *Random Forests*, an ensemble of decision trees, are especially common because of their robustness to noise and interpretability (feature importance measures). Many works report Random Forest classifiers achieving top performance on stress datasets by leveraging a diverse set of physiological features. For instance, *Dahal et al.* (2023) developed a random-forest-based model using a reduced subset of HRV features to detect global stress levels, achieving an efficient and generalizable classifier. In addition to these, researchers have explored **Naïve Bayes, linear discriminant analysis (LDA)**, and **multi-layer perceptrons** for stress classification, often as baseline models in comparisons.

Notably, **ensemble learning techniques** have shown particular effectiveness in stress detection tasks. Ensemble methods combine the predictions of multiple models to improve generalization. A prominent example is the use of **gradient boosted trees**, such as XGBoost, which was identified in a recent survey as one of the most effective ensemble algorithms for wearable stress recognition. *Tazarv et al.* (2021) and *Xefteris et al.* (2023) each employed XGBoost in their stress monitoring frameworks, citing its ability to handle feature interactions and imbalanced data common in stress datasets. These ensemble tree models often outperformed single classifiers like SVM in cross-dataset evaluations. Other works have combined multiple classifier outputs via majority voting or stacking to improve robustness. For example, *Yadav and Bokhari* (2023) conducted a comparative study of mental stress detection algorithms and found that hybrid models (blending SVM, decision tree, and neural network outputs) yielded higher accuracy than any single algorithm alone. Some research has also looked at fusing predictions over time: *Can et al.* (2020) introduced a **decision-level smoothing** mechanism, clustering personal stress levels and smoothing classifier outputs across consecutive time windows to filter out spurious detections. This can be seen as an ensemble over time, improving reliability for practical use. In summary, classical ML techniques – especially ensemble and hybrid models – form a core part of current stress detection systems, with many studies reporting accuracy improvements of a few percentage points by using ensembles (Random Forest, XGBoost) over single-model approaches. The relative simplicity and low computational cost of these models also make them attractive for implementation on wearable or mobile platforms in real-time.

## **Deep Learning and Hybrid Models for Stress Detection**

Alongside classical ML, there has been a surge of interest in **deep learning (DL)** approaches for stress detection in the last few years. Deep learning models can automatically learn complex features from raw sensor data, potentially capturing subtle patterns that manual feature engineering might miss. The most common deep architectures applied are **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)** (particularly Long Short-Term Memory networks, LSTMs). CNNs excel at learning spatial or local temporal patterns and have been used to extract features from physiological time-series (sometimes by transforming signals into spectrograms or images for 2D CNN input). *Zhang et al.* (2021) developed a CNN-based approach for real-time stress detection using raw ECG signals as input, achieving high accuracy in distinguishing stress versus calm states on a public dataset. Recurrent networks like LSTMs are well-suited to sequence modeling and have been employed to capture longer-term temporal dependencies in stress data – for example, changes in HRV or EDA over several minutes. *Mikhaylov et al.* (2022) combined CNN and LSTM layers in a hybrid deep model to leverage CNN’s feature extraction with LSTM’s temporal modeling, demonstrating improved detection of stress episodes during driving tasks. In one noteworthy study, *Xiang et al.* (2025) proposed a **multimodal deep learning framework** that integrates both time-domain and frequency-domain features from multiple signals through a parallel CNN architecture, followed by fully connected layers. Their model, evaluated on a nursing professionals dataset, achieved an accuracy of 91% and F1-score 0.91, significantly outperforming traditional classifiers such as logistic regression, Random Forest, and XGBoost on the same data. This highlights the *potential gains* deep learning can provide when sufficient training data and carefully designed architectures are available.

However, deep learning models also introduce challenges. They typically require large labeled datasets for training, whereas stress datasets are often relatively small (dozens of subjects with a few hours of data each). To address data scarcity, researchers have explored data augmentation and transfer learning. Some have used **generative adversarial networks (GANs)** to generate synthetic physiological data for training; for example, *Moser et al.* (2024) augmented an LSTM classifier with a *Deep Generative Ensemble* of GANs to enrich the training set, which improved stress prediction recall and precision by a few percentage points. Others leverage **pre-training** on related tasks: *Islam and Washington* (2023) employed self-supervised pre-training on unlabeled wearable data to learn generalized representations before fine-tuning a stress classifier, boosting performance on individuals with limited data. Another issue is the **interpretability** of deep models. Black-box neural networks can be difficult to trust in medical or well-being applications. In response, recent works have incorporated explainability techniques. *Moser et al.* (2024) applied Integrated Gradients (an explainable AI method) to their LSTM model to identify which features (e.g., certain EDA peaks or HRV changes) were most influential in the model’s stress predictions, and showed these correspond to known stress biomarkers in the literature. Such efforts aim to validate that deep models are learning meaningful physiological cues rather than spurious patterns. Despite these challenges, the trend is toward **hybrid models** that combine the strengths of deep learning and traditional approaches. For instance, some frameworks use deep learning to automatically extract features, then feed those features into a lighter-weight classifier (or vice versa). The consensus in recent surveys is that deep learning can substantially improve stress detection performance, **provided that issues of data quantity, personalization, and transparency are properly handled**. As datasets grow and sharing of public stress databases (like the popular WESAD dataset) increases, we can expect deep models to play an ever larger role in this field.

## **Datasets, Evaluation Metrics, and Key Findings**

Research on stress detection draws on a mix of public and private datasets, each with different sensor setups and stress elicitation protocols. One widely used benchmark is the **WESAD dataset** (Wearable Stress and Affect Detection), which contains multimodal data (chest and wrist sensors measuring ECG, EDA, EMG, respiration, temperature, accelerometry) from subjects undergoing scripted stress-induction (e.g., the Trier Social Stress Test) and baseline relaxation. WESAD and similar datasets have facilitated comparative studies – many authors report their accuracy, precision, recall, and F1-score on these common datasets to demonstrate improvements. Reported performance varies based on the scenario: in controlled lab settings with clear stress events, classification accuracies in the high 80s to 90% are often achieved. For example, *Zhang et al.* (2021) reported ~93% accuracy using a deep CNN on an ECG-based stress dataset, and *Xiang et al.* (2025) achieved 91% accuracy as noted with their multimodal deep model. In semi-controlled or real-world datasets, performance tends to be lower; achieving 70–80% accuracy is more common due to noise and the difficulty of labeling “ground truth” stress in everyday life. A recent systematic review and meta-analysis by *Abd-alrazaq et al.* (2024) evaluated 19 studies of wearable AI for student stress detection. The meta-analysis found a pooled average accuracy of **85.6%** for stress classification and an average F1-score of ~0.76 across studies. Notably, the review identified several factors that significantly influence performance: the number of stress classes (binary vs multi-level), the type and placement of sensors, the size of the dataset, and the method of ground truth labeling all had measurable effects on accuracy. For instance, models built on larger datasets tended to perform better (dataset size was a significant moderator of accuracy), and studies using objective biochemical markers or well-defined stressor tasks as ground truth sometimes showed higher consistency than those using subjective self-reports. Common evaluation metrics in this domain include accuracy, sensitivity (recall), specificity, and the F1-score, since the class imbalance (much more “non-stress” data than “stress” data typically) can be an issue. Cross-validation and external validation on separate cohorts are used to test generalizability. An important finding across many studies is that **personalization can improve performance** – models tuned to an individual’s baseline and responses often outperform one-size-fits-all models. This is because there are significant inter-personal differences in stress reactivity and sensor signal patterns. Techniques like personal calibration sessions, transfer learning between subjects, or clustering individuals by physiological response profiles (as done by *Can et al.*, 2020) have been used to tailor models more closely to users, yielding more reliable detection in practice. In summary, the literature demonstrates that high accuracies are attainable in recognizing stress from wearables, especially in controlled conditions, but performance can degrade in more naturalistic settings. The *key findings* point to the value of multimodal data, the need for larger and more diverse datasets, and the benefit of personalized or adaptive modeling to account for individual variability.

## **Limitations and Open Challenges**

Despite considerable progress, current stress detection approaches face several **limitations** that hinder generalizability and real-world deployment. A primary concern is the generalizability of models across different users, contexts, and sensor devices. Many studies train and test on a single dataset collected in a specific environment; when these models are applied to new populations or outside the lab, accuracy often drops. *Vos et al.* (2023) in a systematic review highlighted that model performance often does not transfer well between datasets, underscoring the need for algorithms that are robust to inter-individual differences and varying conditions. Relatedly, most published works involve relatively **small sample sizes** – often on the order of tens of participants – which raises concerns about overfitting and the statistical power of reported results. The meta-analysis by Abd-alrazaq *et al.* noted that studies with larger datasets tended to report higher accuracy, and explicitly pointed out dataset size as a significant factor (p = 0.009) affecting outcomes. This suggests that some earlier results on small samples might be overly optimistic and that more extensive data collection is required to truly validate models. Another fundamental challenge lies in obtaining reliable **ground truth labels** for stress. There is no simple “gold standard” for labeling when a person is stressed versus not; researchers have used proxies like self-reported stress levels, expert annotations of observed behavior, or predefined stressor events (public speaking, mental arithmetic, etc.). Each of these has drawbacks: self-reports are subjective and can be inconsistent, and lab stressors may not reflect the diffuse, chronic stresses of daily life. As a result, training data can contain noisy labels, which in turn limits model accuracy. Innovative labeling methods (e.g., leveraging cortisol or other biochemical markers, or multi-rater EMA protocols) are being explored to improve ground truth quality. Furthermore, **practical deployment issues** must be considered. Wearable sensors, while convenient, introduce concerns about user comfort and adherence – if a device is uncomfortable or battery-intensive, users may not wear it continuously. There are also privacy considerations around continuously collecting sensitive physiological and behavioral data; ensuring secure data handling and user consent is paramount in real-world applications. From a technical standpoint, many current models are power-hungry or computationally intensive (especially deep learning ones), which can be problematic for on-device processing on wearables or smartphones. Efforts in model compression and edge computing are needed to bridge this gap. Finally, while many prototypes exist, **few systems have undergone longitudinal real-world trials** to assess their impact on users’ mental health outcomes. Detected stress is only useful if the system can intervene or help the user manage it. Determining how best to integrate these detection systems with feedback or interventions (vibrational alerts, breathing exercises, context-aware suggestions, clinical escalation, etc.) is an open area of research. In summary, current stress detection technologies show great promise but *remain in a nascent stage*. As noted in recent reviews, wearable AI for stress is *“promising but currently has suboptimal performance”* and should complement (not replace) traditional assessment until further improvements are made. Overcoming issues of generalizability, data scarcity, and reliability will be crucial to moving from research prototypes to deployed mental health monitoring solutions. Ongoing work is focusing on larger-scale studies, federated learning across diverse data sources, and combining physiological sensors with behavioral context to improve robustness. Addressing these challenges will pave the way for **deployable stress monitoring systems** that can operate reliably across populations and ultimately provide timely support to individuals in managing their mental well-being.

## **References**

[1] A. Pinge *et al.*, “Detection and monitoring of stress using wearables: a systematic review,” *Frontiers in Computer Science*, vol. 6, article 1478851, Dec. 2024. DOI: 10.3389/fcomp.2024.1478851

[2] J.-Z. Xiang *et al.*, “A multi-modal deep learning approach for stress detection using physiological signals: integrating time and frequency domain features,” *Frontiers in Physiology*, vol. 16, 1584299, Apr. 2025. DOI: 10.3389/fphys.2025.1584299

[3] M. K. Moser, B. Resch, and M. Ehrhart, “An explainable deep learning approach for stress detection in wearable sensor measurements,” *Sensors*, vol. 24, no. 16, p. 5085, 2024. DOI: 10.3390/s24165085

[4] S. Gedam and S. Paul, “A review on mental stress detection using wearable sensors and machine learning techniques,” *IEEE Access*, vol. 9, pp. 84045–84066, 2021. DOI: 10.1109/ACCESS.2021.3085502

[5] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, “Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review,” *Int. J. Med. Informatics*, vol. 173, p. 105026, 2023. DOI: 10.1016/j.ijmedinf.2023.105026

[6] A. Abd-alrazaq *et al.*, “The performance of wearable AI in detecting stress among students: systematic review and meta-analysis,” *J. Med. Internet Res.*, vol. 26, e52622, 2024. DOI: 10.2196/52622

[7] N. Gomes, M. Pato, A. R. Lourenço, and N. Dâmaso, “A survey on wearable sensors for mental health monitoring,” *Sensors*, vol. 23, no. 3, p. 1330, 2023. DOI: 10.3390/s23031330

[8] T. Chalmers *et al.*, “Stress Watch: the use of heart rate and heart rate variability to detect stress – a pilot study using smart watch wearables,” *Sensors*, vol. 22, no. 1, p. 151, 2022. DOI: 10.3390/s22010151

[9] L. Zhu *et al.*, “Stress detection through wrist-based electrodermal activity monitoring and machine learning,” *IEEE J. Biomed. Health Informatics*, vol. 27, no. 5, pp. 2155–2165, 2023. DOI: 10.1109/JBHI.2023.3239305

[10] Y. Cho, S. J. Julier, and N. Bianchi-Berthouze, “Instant stress: detection of perceived mental stress through smartphone photoplethysmography and thermal imaging,” *JMIR Mental Health*, vol. 6, no. 4, e10140, 2019. DOI: 10.2196/10140

[11] Y. S. Can, B. Arnrich, and C. Ersoy, “Stress detection in daily life scenarios using smartphones and wearable sensors: a survey,” *J. Biomed. Informatics*, vol. 92, p. 103139, 2019. DOI: 10.1016/j.jbi.2019.103139

[12] Y. S. Can *et al.*, “Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches,” *IEEE Access*, vol. 8, pp. 38146–38163, 2020. DOI: 10.1109/ACCESS.2020.2975351

[13] S. Campanella *et al.*, “A method for stress detection using Empatica E4 bracelet and machine-learning techniques,” *Sensors*, vol. 23, no. 7, p. 3565, 2023. DOI: 10.3390/s23073565

[14] S. Aristizabal *et al.*, “The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment,” *IEEE Access*, vol. 9, pp. 102053–102068, 2021. DOI: 10.1109/ACCESS.2021.3097038

[15] P. Zhang *et al.*, “Real-time psychological stress detection according to ECG using deep learning,” *Applied Sciences*, vol. 11, no. 9, p. 3838, 2021. DOI: 10.3390/app11093838

[16] K. Dahal, B. Bogue-Jimenez, and A. Doblas, “Global stress detection framework combining a reduced set of HRV features and random forest model,” *Sensors*, vol. 23, no. 11, p. 5220, 2023. DOI: 10.3390/s23115220