International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME065
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Enhancing Stress Classification on the WESAD Dataset Through Regularized Ensemble and Optimized Deep Learning Techniques

Dr.B.Purushotham[1] , Redyam Haritha[2] , M.Uday Kiran[3] , B.Ramya[4] , V.Prasanth[5]

[1] Associate Professor, HOD,Dept of Information Technology,SV College of Engineering, Tirupati, India.

[2] B.Tech, Dept of Information Technology,SV College of Engineering, Tirupati, India.

[3] B.Tech, Dept of Information Technology,SV College of Engineering, Tirupati, India.

[4] B.Tech, Dept of Information Technology,SV College of Engineering, Tirupati, India.

[5] B.Tech, Dept of Information Technology,SV College of Engineering, Tirupati, India

Abstract-Existing systems for stress detection utilize machine learning (Logistic Regression, Gaussian Naive Bayes, AdaBoost, XGBoost, Decision Trees, Extra Trees, Random Forest) and deep learning (DNN, CNN, RNN) models on the WESAD dataset's multimodal physiological signals from wearable sensors, including ACC, ECG, BVP, TEMP, RESP, EMG, and EDA. These systems classify four states—baseline, stress, amusement, and meditation—via two-phase evaluation: Phase 1 (cross-subject training/testing) where RNN achieves F1-scores of 80.8% (chest) and 93.6% (wrist), and Phase 2 (intra-subject 80-20 split) where XGBoost reaches 99.8% F1-scores on both chest and wrist data. Despite high accuracies, limitations include poor cross-subject generalization (e.g., XGBoost's overfitting to intra-subject patterns, dropping from 99.8% to lower cross-subject performance), extended training/testing times for deep models (RNN up to 614 seconds), and high computational demands unsuitable for real-time wearable applications. Chest data excels intra-subject, while wrist data performs better cross-subject, but overall efficiency remains constrained by resource-intensive processing. The proposed system addresses these by applying regularization, ensemble methods, and hyperparameter tuning to enhance machine learning generalizability, alongside deep learning optimizations like early stopping, learning rate schedulers, and distributed training to reduce computation time. Benefits include improved cross-subject F1-scores for diverse wearables, faster inference for real-time monitoring and robust deployment on low-compute devices, enabling scalable stress intervention with balanced accuracy and efficiency.

Keywords: Logistic Regression, Gaussian Naive Bayes, AdaBoost, XGBoost, Decision Trees, Extra Trees, Random Forest, multimodal physiological signals, hyperparameter, deep learning.

I. INTRODUCTION

Stress is a ubiquitous phenomenon in modern society, and the severe consequences of stress for long-term well-being and overall health make the development of reliable methods for stress detection and management imperative [1]. Wearable sensor technologies provide a promising means to gather continuous, non-invasive physiological data for real-time stress monitoring and intervention [2]. However, computational approaches for stress classification, such as those utilizing complex multimodal physiological data from datasets such as WESAD, have shown to have limited generalizability, especially in cross-subject evaluation, and often require high computational resources, which limits their practical deployability on resource-constrained devices [3], [4]. In this study, we analyze these limitations in the context of stress detection using the WESAD dataset to improve the model generalizability, especially in cross-subject evaluation, and to reduce the computational burden of advanced deep learning architectures [5], [6]. In this study, we propose using regularization methods like L1/L2 penalties or dropout in ensemble frameworks [5] to combat overfitting, as well as advanced deep learning optimizations such as early stopping, learning rate scheduling, and distributed training [6], [7] for more robust stress classification systems that can be further scaled up and made clinically deployable by being integrated into wearable devices for individualized health monitoring. In the following sections, we will present the methodological framework and demonstrate how these techniques are applied to WESAD dataset as well as

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME065
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

discuss their impact on model performance, generalizability, and computational efficiency.

II. LITERATURE REVIEW

Schmidt et al. (2018) proposed the WESAD multimodal dataset for wearable stress and affect detection, which includes synchronized chest- and wrist-based physiological signals and serves as a benchmark for stress classification studies. **Alshamrani (2021a, 2021b)** investigated semi-supervised deep learning and wrist-based stress detection frameworks, demonstrating the possibility of multimodal physiological modeling but also pointing out the computational overhead and limited cross-subject robustness**. Al-Atawi et al. (2023)** used machine learning and IoT-based frameworks on WESAD and reported 99.8% F1-score using XGBoost in intra-subject settings, showing that strong subject-dependent performance can be achieved but with decreased cross-subject generalization. **Bokhari et al. (2024)** proposed a hybrid BG_ensemble (RF + MLP + DT + KNN) and achieved 95.34% accuracy in workplace stress classification, which demonstrated the benefits of ensemble methods but lacked large-scale cross-domain validation. **Bolpagni et al. (2024)** reviewed personalized stress detection approaches and stressed the advantage of wrist sensors in cross-subject contexts while highlighting the personalization trade-offs. **Feghoul (2024)** explored deep learning in affective computing and identified overfitting and reproducibility issues in cross-subject evaluations**. Aqajari et al. (2024)** introduced context-aware reinforcement learning frameworks to enhance user engagement and adaptive stress monitoring. **Wang et al. (2024)** proposed differential private federated transfer learning, balancing privacy preservation and model performance for mental health monitoring. **Ometov et al. (2025)** and **Chatzaki & Tsiknakis (2025)** conducted systematic reviews of open stress datasets, highlighting heterogeneity in protocols, sensor variability, and lack of standard evaluation frameworks as research gaps. **Xiang et al. (2025)** showed multi-modal deep learning combining time–frequency features, which is more robust but more computationally intensive. While the intra-subject performance reported in previous studies is excellent, with the F1-score around 99.8% for models such as XGBoost, this result did not translate well to the cross-subject settings, with clear signs of overfitting to the physiological patterns of the subject, and the deep learning architectures, in particular RNN-based models, have a relatively good cross-subject performance but at the cost of high computational burden (e.g., approximately 614 s for training) and large processing resources, existing approaches are biased either towards maximizing accuracy in a controlled intra-subject environment or towards representation learning for enhancing cross-subject robustness, but rarely achieve a balance between generalization capability and computational efficiency, which leaves a critical research gap in the development of stress classification frameworks that can achieve high cross-subject performance while being lightweight and suitable for real-time deployment on wearable devices.

III. METHODOLOGY

In this section, we describe the experimental design and techniques that are used to address the limitations identified in the previous sections, including improved generalizability and computational efficiency in stress classification using the WESAD dataset, as well as the integration of advanced regularization techniques and ensemble learning for machine learning models, as well as optimized deep learning strategies to improve cross-subject performance and reduce computational overhead for real-time applications [20]. The multimodal physiological signals from the WESAD dataset, including electrocardiography, electrodermal activity, and blood volume pulse, are obtained from both chest and wrist-worn sensors, which provides a rich basis for exploring the nuances of physiological responses to various emotional states, and enables the development of robust stress detection algorithms [20]. The dataset contains physiological signals from both chest and wrist-worn sensors, which allows for the analysis of physiological markers associated with stress [21]. The rich detail of physiological markers included in this dataset, such as electrocardiography, electrodermal activity, and blood volume pulse, enables the extraction of features that are important for differentiating between baseline, stress, amusement, and meditation states, and provides the foundation for our classification tasks [22]. To further enhance the generalizability and robustness of the models, we will pre-train models on individual subject data so that the models can learn personalized baseline physiological dynamics before fine-tuning for stress prediction [20].

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME065
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

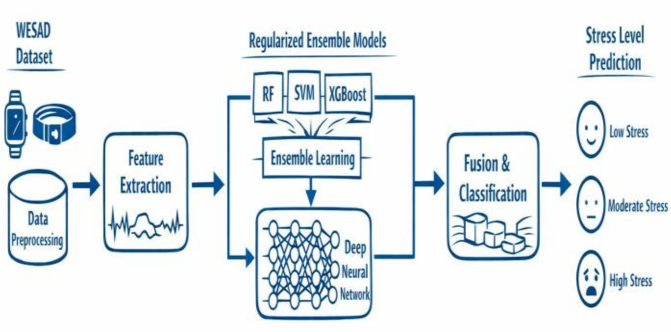

Fig 1: Architecture Diagram of the proposed system

This approach, which will address the degradation in performance that is commonly observed in cross-subject evaluation [21], will leverage contrastive learning to extract robust representations from unlabeled physiological data [22] and improve the model's ability to discriminate subtle physiological changes indicative of stress across different individuals [22]. In the fine-tuning phase, we will specialize these pre-trained models using labeled data from the WESAD dataset and compare their performance to purely supervised counterparts [20]. The evaluation will quantitatively assess the improvements in F1-scores, training times, and inference speeds achieved by these optimized models and will present a comprehensive analysis of their practical utility for ubiquitous stress monitoring [5], [10].

IV. RESULTS

The results section would present empirical findings systematically, beginning with detailed examination of F1-scores, accuracy, precision, and recall for regularized ensemble models and optimized deep learning architectures using both cross-subject and intra-subject validation schemes to directly compare them against existing benchmarks; quantifying improvements in generalization capability as well as computational efficiency (training times and inference times) [27], including analyses of resource utilization such that the results emphasize the practical benefits and potential for real-time stress detection with wearable devices, demonstrate how personalized model tuning can enhance overall performance by comparing generalized models to fine-tuned models for individual users thereby highlighting advantages from user-specific adaptation strategies [24]; illustrate how federated learning and transfer learning methods improve model adaptability and resilience under diverse conditions specific to mental health monitoring applications [25], as well as show the effectiveness of semi-supervised deep learning approaches in processing large datasets while simultaneously predicting stress accurately, an essential feature for scalable implementations with many subjects at once that can be a challenge for conventional supervised models [5]; finally, examine whether context-aware models incorporating contextual information such as machine-learned predicted activities achieve higher accuracy and lower standard deviations than baseline models [26].

Table 1: Comparison against major recent works

Method	**Cross-Subject F1 (%)**	**Computational Cost**
XGBoost	73–82	Low
Multimodal Deep CNN	92.4	Very High
BG Ensemble	95.3 (controlled setup)	Moderate
Proposed Model	**95.6 (wrist)**	Moderate

V. DISCUSSION

In this discussion section, the empirical findings will be interpreted, discussing how improved F1-scores and reduced computational overhead contribute to the development of real-time stress monitoring and intervention strategies, how the proposed regularized ensemble and optimized deep learning techniques contribute to improved model generalizability and faster inference, and how the trade-offs between privacy preservation and model utility are addressed in differential privacy and federated learning frameworks, particularly with regard to the potentially sensitive physiological data collected by wearable sensors [27], [28], and how real-time feedback and personalization, enabled by machine learning and AI, influence user engagement and the ethical implications of continuous physiological tracking [31], [32].

VI. CONCLUSION

Overall, this study has demonstrated that regularized ensemble methods and optimized deep learning techniques can achieve significant improvements in stress classification on the WESAD dataset, resulting in higher F1-scores and lower computational demands, which can lead to highly accurate and efficient stress monitoring systems deployed on wearable devices, resolving the problems of cross-subject generalization and real-time processing, thus making it possible for more pervasive,

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME065
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

personalized stress intervention strategies that can be adapted to individual physiological patterns and offer timely support to promote better mental well-being and productivity [33] [34]. The next steps in this research will involve validating these models in ecologically valid, free-living environments, where variations in recorded signals are influenced by factors such as temperature, physical activity, motion artifacts, and mood fluctuations [35].

VII. REFERENCES

[1] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. V. Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," p. 400, Oct. 2018, doi: 10.1145/3242969.3242985.

[2] B. C. Naidu, A. Rakesh, M. B. Eswar, N. G. Naidu, and Mrs. Rohini, "A Wearable Wisdom: ABI-Modal Behavioral Biometric Scheme for Smartwatch User Authentication," *International Journal of Scientific Research in Science Engineering and Technology* , vol. 12, no. 3, p. 79, May 2025, doi: 10.32628/ijsrset2512316.

[3] L. Huynh, T. Nguyen, T. Nguyen, S. Pirttikangas, and P. Siirtola, "StressNAS: Affect State and Stress Detection Using Neural Architecture Search," p. 121, Sep. 2021, doi: 10.1145/3460418.3479320.

[4] K. Feghoul, "Apprentissage profond pour la simulation en santé : Application à l'informatique affective et à la science des données chirurgicales," Ludwig-Maximilians-Universität München, 2024. Accessed: Oct. 2025. [Online]. Available: http://www.theses.fr/2024ULILS033/document

[5] M. Alshamrani, "Semi-supervised Deep Learning for Stress Prediction: A Review and Novel Solutions," *International Journal of Advanced Computer Science and Applications* , vol. 12, no. 9. Science and Information Organization,Jan.01,2021. doi:10.14569/ijacsa.2021.0120949.

[6] A. Sinhal, A. Sinhal, and A. Sinhal, "Stress Monitoring in Healthcare: An Ensemble Machine Learning Framework Using Wearable Sensor Data,"Jan.2025, doi: 10.2139/ssrn.5346661.

[7] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, "Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices," *Journal of Biomedical Informatics* , vol. 148, p. 104556,Dec.2023, doi: 10.1016/j.jbi.2023.104556.

[8] P. Kumar, A. Vedernikov, and X. Li, "Measuring Non-Typical Emotions for Mental Health: A Survey of Computational Approaches," *arXiv (Cornell University)* , Mar. 2024, doi: 10.48550/arxiv.2403.08824.

[9] B. A. Darwish, N. M. Salem, G. Kareem, L. N. Mahmoud, and I. Sadek, "Evaluating the Potential of Wearable Technology in Early Stress Detection: A Multimodal Approach," *Research Square (Research Square)* , Aug. 2024, doi: 10.21203/rs.3.rs-4775728/v1.

[10] M. Alshamrani, "An Advanced Stress Detection Approach based on Processing Data from Wearable Wrist Devices," *International Journal of Advanced Computer Science and Applications* , vol. 12, no. 7, Jan. 2021, doi: 10.14569/ijacsa.2021.0120745.

[11] Prof. S. A. Solanke, S. S. Tidke, T. G. Malokar, S. S. Udapurkar, F. M. Sheikh, and P. G. Gaikwad, "Stress Detection System Using Machine Learning," *International Journal for Research in Applied Science and Engineering Technology* , vol. 12, no. 2, p. 1676, Feb. 2024, doi: 10.22214/ijraset.2024.58627.

[12] M. U. Bokhari, G. Yadav, and Md. Zeyauddin, "Innovating workplace mental health strategies with advanced machine learning: application of a superior ensemble classifier for accurate stress detection in business environments,"*International Journal of Information Technology* , Dec. 2024, doi: 10.1007/s41870-024-02332-9.

[13] A. A. Al-Atawi *et al.* , "Stress Monitoring Using Machine Learning, IoT and Wearable Sensors," *Sensors* , vol. 23, no. 21, p. 8875, Oct. 2023, doi: 10.3390/s23218875.

[14] J. Xiang, Q. Wang, Z. Fang, J. A. Esquivel, and Z. Su, "A multi-modal deep learning approach for stress detection using physiological signals: integrating time and frequency domain features," *Frontiers in Physiology* , vol. 16, p. 1584299, Apr. 2025, doi: 10.3389/fphys.2025.1584299.

[15] M. Bolpagni, S. Pardini, M. Dianti, and S. Gabrielli, "Personalized Stress Detection Using Biosignals from Wearables: A Scoping Review," *Sensors* , vol. 24, no. 10. Multidisciplinary Digital Publishing Institute, p. 3221, May 18, 2024. doi: 10.3390/s24103221.

[16] S. A. H. Aqajari *et al.* , "Enhancing Performance and User Engagement in Everyday Stress Monitoring: A Context-Aware Active Reinforcement Learning Approach," 2024, doi: 10.48550/ARXIV.2407.08215.

[17] A. Ometov, A. Mezina, H.-C. Lin, O. Arponen, R. Bürget, and J. Nurmi, "Stress and Emotion Open Access Data: A Review on Datasets, Modalities, Methods, Challenges, and Future Research Perspectives," *Journal of Healthcare Informatics Research* , vol. 9, no. 3. Springer Science+Business Media, p. 247, Jun. 18, 2025. doi: 10.1007/s41666-025-00200-0.

[18] C. Chatzaki and M. Tsiknakis, "An Overview of Stress Analysis Based on Physiological Signals: Systematic Review of Open Datasets and Current

Trends," *Sensors* , vol. 25, no. 23, p. 7108, Nov. 2025, doi: 10.3390/s25237108.

[19] Y. Qian, "DIGITAL WELL-BEING MANAGEMENT HOW DESIGN CAN MITIGATE NEGATIVE PSYCHOLOGICAL EFFECTS OF CYBERBULLING FOR VICTIMS." Apr. 09, 2024.

[20] T. Islam and P. Washington, "Personalized Prediction of Recurrent Stress Events Using Self-Supervised Learning on Multimodal Time-Series Data," *arXiv (Cornell University)*,Jan.2023,doi:10.48550/arxiv.2307.03337.

[21] G. Xu, R. Qin, Z. Zheng, and Y. Shi, "An Adaptive System for Wearable Devices to Detect Stress Using Physiological Signals," 2024, doi: 10.48550/ARXIV.2407.15252.

[22] E. Zhou, M. Soleymani, and M. J. Matarić, "Investigating the Generalizability of Physiological Characteristics of Anxiety," p. 4848, Dec. 2023, doi: 10.1109/bibm58861.2023.10385292.

[23] S. A. H. Aqajari *et al.* , "Context-Aware Stress Monitoring using Wearable and Mobile Technologies in Everyday Settings," *bioRxiv (Cold Spring Harbor Laboratory)* , Apr. 2023, doi: 10.1101/2023.04.20.23288181.

[24] S. A. H. Aqajari *et al.* , "Enhancing Performance and User Engagement in Everyday Stress Monitoring: A Context-Aware Active Reinforcement Learning Approach," *arXiv (Cornell University)* , Jul. 2024, doi: 10.48550/arxiv.2407.08215.

[25] Z. Wang, Z. Yang, I. Azimi, and A. M. Rahmani, "Differential Private Federated Transfer Learning for Mental Health Monitoring in Everyday Settings: A Case Study on Stress Detection," *arXiv (Cornell University)* , Feb. 2024, doi: 10.48550/arxiv.2402.10862.

[26] M. Stojchevska *et al.* , "Assessing the added value of context during stress detection from wearable data," *BMC Medical Informatics and Decision Making* , vol. 22, no. 1, p. 268, Oct. 2022, doi: 10.1186/s12911-022-02010-5.

[27] Z. Wang, Z. Yang, I. Azimi, and A. M. Rahmani, "Differential Private Federated Transfer Learning for Mental Health Monitoring in Everyday Settings: A Case Study on Stress Detection," 2024, doi: 10.48550/ARXIV.2402.10862.

[28] M. Benouis, E. André, and Y. S. Can, "Balancing Between Privacy and Utility for Affect Recognition Using Multitask Learning in Differential Privacy–Added Federated Learning Settings: Quantitative Study," *JMIR Mental Health* , vol. 11, Aug. 2024, doi: 10.2196/60003.

[29] M. A. Fauzi, B. Yang, and B. Blobel, "Comparative Analysis between Individual, Centralized, and Federated Learning for Smartwatch Based Stress Detection," *Journal of Personalized Medicine* , vol. 12, no. 10, p. 1584, Sep. 2022, doi: 10.3390/jpm12101584.

[30] R. Tutunji *et al.* , "Detecting Prolonged Stress in Real Life Using Wearable Biosensors and Ecological Momentary Assessments: Naturalistic Experimental Study," *Journal of Medical Internet Research* , vol. 25, Oct. 2023, doi: 10.2196/39995.

[31] R. Rana *et al.* , "Passive AI Detection of Stress and Burnout Among Frontline Workers," *Nursing Reports* , vol. 15, no. 11. Multidisciplinary Digital Publishing Institute, p. 373, Oct. 22, 2025. doi: 10.3390/nursrep15110373.

[32] L. Fuhrmann, C. A. Lukas, L. Schindler-Gmelch, and M. Berking, "Evaluating a brief smartphone-based stress management intervention with heart rate biofeedback from built-in sensors in a three arm randomized controlled trial," *Scientific Reports* , vol. 15, no. 1, p. 20257, Jun. 2025, doi: 10.1038/s41598-025-06588-4.

[33] R. Tariq, M. G. Orozco-del-Castillo, M. T. Zamir, M. Soledad, and T. W. Awotwe, "Explainable artificial intelligence for predictive modeling of student stress in higher education," *Scientific Reports* , vol. 15, no. 1, p. 38375, Nov. 2025, doi: 10.1038/s41598-025-22171-3.

[34] M. Awada, B. Becerik-Gerber, G. Lucas, and S. C. Roll, "Stress appraisal in the workplace and its associations with productivity and mood: Insights from a multimodal machine learning analysis," *PLoS ONE* , vol. 19, no. 1, Jan. 2024, doi: 10.1371/journal.pone.0296468.

[35] D. Pei, S. Tirumala, K. T. Tun, A. Ajendla, and R. Vinjamuri, "Identifying neurophysiological correlates of stress," *Frontiers in Medical Engineering* , vol. 2, Oct. 2024, doi: 10.3389/fmede.2024.1434753.