A Dual-Branch Convolutional Neural Network with Gated Recurrent Units Network for Enhanced Multimodal Stress Monitoring from Wearable Physiological Signals

Md Rokonuzzaman Mim
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
rokonuzzaman@vu.edu.bd

Md Santo Ali
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
shantorueteee19@ieee.org

Mohammod Abdul Motin
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
m.a.motin@ieee.org

Mufti Mahmud
Department of Information & Computer Science
King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia
mufti.mahmud@kfupm.edu.sa

Abstract

Chronic mental stress poses severe threats to both physical and psychological well being, highlighting the importance of continuous monitoring through wearable technologies. Blood Volume Pulse (BVP) and Electrodermal Activity (EDA) signals provide reliable, noninvasive, and cost-effective means for stress assessment. In this work, we present a lightweight deep learning framework that integrates dual-branch convolutional neural networks (CNN) with gated recurrent units (GRU) for real-time stress detection from multimodal BVP and EDA signals. The model is evaluated on the publicly available WESAD dataset using subject-independent leave-one-subject-out validation and achieves state of the art performance: 99.27% accuracy, 99.97% F1-score, 99.68% AUC, and 98.40% Cohen's κ. To address class imbalance, a sliding window augmentation strategy is employed, significantly boosting the minority class performance. With only 0.43M parameters and minimal computational cost, the proposed architecture is optimized for deployment on resource-constrained wearable devices, offering a robust solution for real-world stress monitoring.

1 Introduction

Mental stress negatively impacts both physical and mental health, contributing to cardiovascular disease, hypertension, depression, cognitive decline, and immune dysfunction [1, 2]. These risks highlight the need for reliable, continuous stress monitoring for clinical and everyday well being. Commonly used physiological signals for stress detection include electrocardiography (ECG), electroencephalography (EEG), electrodermal activity (EDA) and blood volume pulse (BVP) [3–5]. BVP reflects peripheral blood flow and heart rate variability modulated by both sympathetic and parasym-

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

pathetic activity [3, 6, 7], while EDA captures skin conductance changes driven solely by sympathetic activation, serving as a direct indicator of emotional and physiological arousal [4, 5, 8]. Together, these signals offer complementary insights into autonomic dynamics for reliable and computationally efficient stress estimation.

Although signals such as ECG, EMG, EEG, and ACC offer valuable insights into stress-related physiological responses, their application in continuous wearable monitoring is limited by multi-electrode setups, motion artifacts, and high power demands [9–12]. In contrast, BVP and EDA can be unobtrusively captured from wrist-worn devices such as Empatica E4 or Fitbit sensors, offering an optimal balance between signal quality and wearability. In recent years, numerous studies have employed machine learning (ML) and deep learning (DL) techniques on various wearable signals for real-time stress monitoring [13–15]. Several studies have explored ML-based methods for stress monitoring [16–19] for stress monitoring. These methods typically rely on handcrafted features such as time, frequency, and nonlinear domains of physiological signals. While these approaches have shown promising performance, their dependence on manually designed, dataset specific features limits adaptability and generalization across subjects and real world scenarios, as handcrafted features often fail to capture complex, nonlinear, and individual-specific patterns in the data. Deep

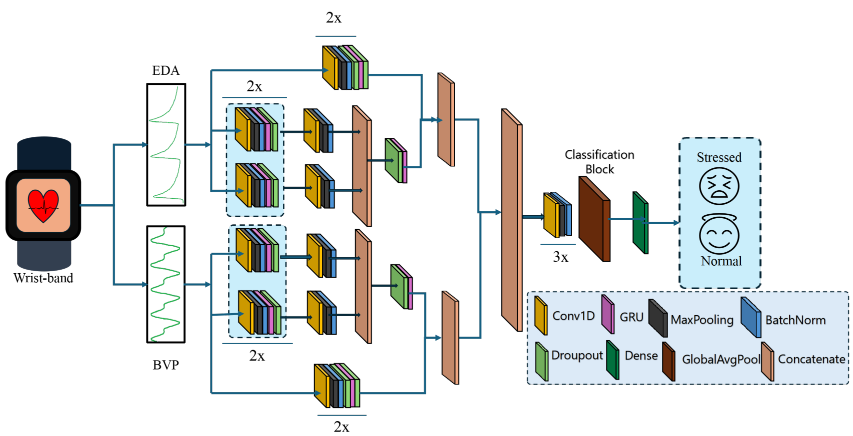

Figure 1: Proposed lightweight CNN-GRU architecture for stress detection using multimodal physiological signals. The model processes simulated normal signals through parallel convolutional and recurrent pathways, incorporating Conv1D, MaxPooling1D, BatchNormalization, GRU, and Dense layers. Feature representations are integrated via concatenation and global average pooling for final classification.

learning (DL) has recently emerged as a powerful approach for stress monitoring due to its ability to automatically extract meaningful representations directly from raw physiological signals such as BVP and EDA [4, 7, 10, 20]. This capability enables DL models to capture complex patterns essential for accurately characterizing stress responses. In contrast, previous studies have largely relied on hand crafted features or computationally intensive transformations of time-series signals into image representations, both of which pose limitations for real-time applications on resource-constrained wearable devices. These challenges underscore the promise of DL based methods for efficient stress detection, unlike ML based approaches. In this work, we present a novel lightweight deep learning architecture combining dual branch convolutional neural network with gated recurrent units (CNN–GRU), specifically designed for enhanced stress monitoring using wearable-derived BVP and EDA signals on resource-constrained devices. To address class imbalance, a sliding-window-based augmentation strategy is employed, significantly improving the recognition of minority classes.

2 Materials and Methods

2.1 Dataset Description

This study utilizes the publicly available Wearable Stress and Affect Detection (WESAD) dataset [21] to evaluate the proposed model. The dataset comprises physiological signals including BVP, ECG, EDA, EMG, and TEMP collected from 17 participants. Due to sensor malfunctions, two participants were excluded, resulting in a final cohort of 15 subjects (12 males and 3 females; mean age 27.5 $\pm$ 2.4 years). For the present analysis, only BVP and EDA signals are utilized, sampled at 64 Hz and 4 Hz, respectively, providing key insights into cardiovascular and autonomic responses relevant to stress detection.

2.2 Signal Preprocessing and Augmentation

The raw BVP and EDA signals were segmented into non-overlapping 30-second windows. Z-score normalization was applied to minimize intersubject amplitude variations. To address class imbalance, sliding window data augmentation was introduced, generating additional samples from minority classes to improve training stability and generalization.

2.3 Proposed Deep Learning Model

The proposed architecture, illustrated in Figure 1, a dual-branch multimodal framework that integrates BVP and EDA signals for stress classification. Each branch begins with stacked Conv1D and MaxPooling1D layers, combined with batch normalization and dropout, to extract stable and noise-resilient features. The resulting feature maps are then passed through GRUs, enabling the model to capture temporal dependencies in both forward and backward directions. The modality-specific outputs are concatenated into a unified representation, further refined through an additional convolutional block with pooling and normalization. Finally, global average pooling compresses the high-level features, and a dense layer with softmax activation performs the classification. By combining convolutional feature extraction with sequential modeling, the framework provides an efficient and robust solution for multimodal stress detection.
The proposed model was evaluated using leave-one-subject-out (LOSO) cross-validation to ensure subject-independent performance, with metrics including accuracy, F1-score, specificity, sensitivity, AUC, and Cohen's kappa (κ).

3 Results and Discussion

The proposed model achieved an accuracy of 97. 53%, a specificity of 98. 87%, a sensitivity of 95. 08%, an F1 score of 96. 14%, an AUC of 98. 4%, and Cohen's κ of 94. 34% without data augmentation. However, due to class imbalance—where normal samples significantly outnumbered stressed samples—the model's sensitivity to the minority (stressed) class was limited. With augmentation, the model's performance improved significantly. Accuracy increased by 1.74% and reached 99.27%, while specificity improved by 0.38% to 99.25%. Sensitivity showed the most notable gain, rising by 4.21% to 99.29%. The F1-score increased by 3.83%, achieving 99.97%, and the AUC rose by 1.21% to 99.68%. Similarly, Cohen's κ improved by 4.06%, reaching 98.40%. These improvements highlight the effectiveness of augmentation in balancing class-wise performance, particularly enhancing detection of the stressed class while maintaining high specificity for the normal class. Furthermore, the AUC improves from 0.98 to 0.99, confirming the robustness of the proposed approach. Detailed performance metrics are provided in Table 1.

Table 1: The performance scores using the leave-one-subject-out strategy, both with and without augmentation, are presented. The accuracy, F1 score, specificity, sensitivity, AUC, and κ are expressed in percentages (%).

Aug	**Acc**	**Spe**	**Sen**	**F1**	**AUC**	κ
No	97.53	98.87	95.08	96.14	98.47	94.34
Yes	99.27	99.25	99.29	99.97	99.68	98.40

3.1 Ablation Study

Moreover, incorporating sliding-window-based augmentation ensured class balance, which contributed to the consistent performance gains observed across all modalities. The ablation results in Table 2 show that EDA alone achieves good performance, with an accuracy of 93.55%. BVP proves even more informative, achieving 97.37% accuracy and an AUC of 98.93%. Moreover, incorporating sliding-window-based augmentation ensured class balance, which contributed to the consistent performance gains observed across all modalities. When both signals are combined, performance improves further, reaching 99.27% accuracy and a κ of 98.40%. These findings indicate that while BVP captures strong temporal information and EDA reflects sympathetic activity, their combination provides the most robust stress classification.

Table 2: Ablation study results showing the performance of individual signals (EDA and BVP) in stress classification. Accuracy (Acc), specificity (Spe), sensitivity (Sen), F1-score (F1), AUC, and Cohen's κ are reported in percentages (%).

Signal	Acc	Spe	Sen	F1	AUC	κ
EDA	93.55	94.34	92.05	90.47	95.92	85.61
BVP	97.37	98.11	96.21	96.01	98.93	94.05
BVP+EDA	99.27	99.25	99.29	99.97	99.68	98.40

3.2 Comparison with Existing Works

A comparative analysis with existing studies was conducted to benchmark the performance of the proposed approach. The closest reported performance achieved 97.90% accuracy [7], which relied solely on unimodal BVP signals, thereby limiting generalizability and practical applicability. In contrast, our proposed method consistently outperforms prior works across, as illustrated in Table 3. The superiority of our approach stems from its ability to effectively handle class imbalance and maintain a lightweight architecture suitable for deployment on resource-constrained devices, thereby demonstrating both higher predictive performance and real-world applicability.

Table 3: Summary of the research study with the existing literature.

Study	Dataset	Signals	Accuracy
[22]	WESAD	BVP	85.00%
[23]	WESAD	BVP	84.17%
[24]	WESAD	BVP	88.56%
[21]	WESAD	BVP, EDA, ACC, TEMP	87.12%
[25]	WESAD	EDA, EEG, PPG	87.40%
[7]	WESAD	BVP	97.90%
[26]	WESAD	BVP	98.23%
Proposed	**WESAD**	**BVP, EDA**	**99.27%**

3.3 Compatibility with Wearable Devices

Continuous stress monitoring is essential for maintaining physical and mental well-being. BVP and EDA sensors are particularly suitable due to their cost-effectiveness, non-invasiveness, and wide adoption in consumer devices [26]. However, wearable platforms are constrained by limited memory and computational resources. The proposed dual-branch CNN-GRU architecture addresses these limitations with only 0.43M parameters (1.64 MB), and 0.068405 GFLOPs, enabling real-time stress monitoring on resource-constrained devices.

4 Conclusions

This study presents a lightweight dual-branch CNN-GRU architecture for multimodal stress detection using BVP and EDA signals from wearable sensors. The model demonstrated exceptional performance on the WESAD dataset, achieving 99.27% accuracy and a κ of 98.40%, outperforming existing methods. A critical factor contributing to this success was the sliding-window augmentation strategy, which effectively mitigated class imbalance and significantly enhanced sensitivity for the minority class. With only 0.43M parameters and low computational overhead, the architecture is highly suitable for real-time deployment on resource-constrained devices. These findings underscore the framework's potential for continuous, non-invasive stress monitoring in everyday settings. Future work will explore multiclass stress detection, cross-dataset validation, and implementation on embedded hardware platforms.

References

[1] M. Esler, "Mental stress and human cardiovascular disease," *Neuroscience & Biobehavioral Reviews*, vol. 74, pp. 269–276, 2017.

[2] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 51–57, IEEE, 2020.

[3] G. Lu, F. Yang, J. A. Taylor, and J. F. Stein, "A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects," *Journal of Medical Engineering & Technology*, vol. 33, no. 8, pp. 634–641, 2009.

[4] S. Elzeiny and M. Qaraqe, "Stress classification using photoplethysmogram-based spatial and frequency domain images," *Sensors*, vol. 20, no. 18, p. 5312, 2020.

[5] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 2018 ACM International Conference on Multimodal Interaction*, pp. 400–408, ACM, 2018.

[6] M. A. Motin, C. K. Karmakar, M. Palaniswami, and T. Penzel, "Photoplethysmographic-based automated sleep–wake classification using a support vector machine," *Physiological Measurement*, vol. 41, no. 7, p. 075013, 2020.

[7] M. S. Ali, M. A. Motin, and M. Mahmud, "A dual path hybrid convolutional neural network and bidirectional long-short term memory approach for ppg-based stress monitoring," in *4th Muslims in ML Workshop co-located with ICML 2025*.

[8] M. Nardelli, A. Greco, G. Valenza, A. Lanata, and E. P. Scilingo, "A wearable system for stress detection through physiological signals analysis," *Biomedical Engineering Online*, vol. 14, no. 1, pp. 1–14, 2015.

[9] S. Pinge, N. Jadhav, and V. Sharma, "Stress detection using multimodal physiological signals: A review of wearable technologies and algorithms," *IEEE Sensors Journal*, 2024. Early Access.

[10] F. Gasparini, A. Grossi, and S. Bandini, "A deep learning approach to recognize cognitive load using ppg signals," in *Proc. PETRA*, pp. 489–495, ACM, 2021.

[11] L. Hongn, J. Zhao, and X. Wang, "A multimodal physiological dataset for stress and exercise analysis with wearable sensors," in *IEEE Transactions on Affective Computing*, 2025. In Press.

[12] M. S. Ali, M. A. Motin, E. S. M. El-Alfy, and M. Mahmud, "A hybrid convolutional neural network-bidirectional long short-term memory approach for PPG-based stress monitoring from wrist worn wearables," in *Neural Information Processing. ICONIP 2025* (T. Taniguchi *et al.*, eds.), vol. 16312 of *Lecture Notes in Computer Science*, (Singapore), Springer, 2026.

[13] P. Kalra and V. Sharma, "Mental stress assessment using ppg signal: A deep neural network approach," *IETE Journal of Research*, vol. 69, no. 2, pp. 879–885, 2023.

[14] R. K. Nath, H. Thapliyal, and A. Caban-Holt, "Machine learning based stress monitoring in older adults using wearable sensors and cortisol as stress biomarker," *Journal of Signal Processing Systems*, vol. 94, no. 6, pp. 513–525, 2022.

[15] M. S. Ali, S. S. Bipro, M. A. Motin, S. Kabir, M. Sharma, and M. Chowdhury, "Teanet: A transpose-enhanced autoencoder network for wearable stress monitoring," *arXiv preprint arXiv:2503.12657*, 2025.

[16] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. ICIRCA*, pp. 1–6, IEEE, 2020.

[17] A. Bellante *et al.*, "Emocy: Towards physiological signals based stress detection," in *Proc. BHI*, pp. 1–4, IEEE, 2021.

[18] S. Heo, S. Kwon, and J. Lee, "Stress detection with single ppg sensor by orchestrating multiple denoising and peak detecting methods," *IEEE Access*, vol. 9, pp. 47777–47785, 2021.

[19] A. Jahanjoo, N. TaheriNejad, and A. Aminifar, “High accuracy stress detection using wrist-worn ppg sensors,” in *Proc. ISCAS*, pp. 1–5, IEEE, 2024.

[20] N. Rashid *et al.*, “Feature augmented hybrid cnn for stress recognition using wrist-based photoplethysmography sensor,” in *Proc. EMBC*, pp. 1–4, IEEE, 2021.

[21] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing wesad, a multimodal dataset for wearable stress and affect detection,” in *Proceedings of the 20th ACM international conference on multimodal interaction*, pp. 400–408, 2018.

[22] M. Alshamrani, “An advanced stress detection approach based on processing data from wearable wrist devices,” *Int. J. Adv. Comput. Sci. Appl*, vol. 12, no. 7, 2021.

[23] P. Garg, J. Santhosh, A. Dengel, and S. Ishimaru, “Stress detection by machine learning and wearable sensors,” in *Companion Proceedings of the 26th International Conference on Intelligent User Interfaces*, pp. 43–45, 2021.

[24] M. Rashid *et al.*, “Stress detection from physiological signals using deep learning,” *IEEE Transactions on Affective Computing*, 2021.

[25] P. Siirtola, “Continuous stress detection using the sensors of commercial smartwatch,” in *Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers*, pp. 1198–1201, 2019.

[26] M. S. Ali, M. A. Motin, and S. Kabir, “Retenet: A residual encoder and transformer encoders network for stress monitoring from wearable device,” *IEEE Signal Processing Letters*, 2025.