A Dual Path Hybrid Convolutional Neural Network and Bidirectional Long-Short Term Memory Approach for PPG-Based Stress Monitoring

Md Santo Ali[1] **Mohammod Abdul Motin**[1] **Mufti Mahmud**[2]

Abstract

Mental stress adversely impacts both physical and mental health, with chronic stress leading to serious health concerns. Photoplethysmography (PPG) sensors, widely available in wearable devices, offer a convenient, cost-effective, and non-invasive method for stress monitoring. This study proposes a dual path hybrid convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM) hybrid architecture for real-time stress detection using only PPG signals. Trained and validated on the publicly available WESAD dataset, the model achieves exceptional performance metrics: 97.90% accuracy, 98.30% specificity, 97.20% sensitivity, 97.06% F1-score, 99.12% AUC, and 95.42% Cohen's kappa. The lightweight model exhibits high accuracy in stress detection while maintaining computational efficiency, making it particularly suitable for wearable devices. These results highlight the potential of this approach for practical real-time stress monitoring and management applications.

1. Introduction

Stress is a crucial adaptive response that helps the body manage challenges and restore balance (Elzeiny & Qaraqe, 2020). However, mental stress has been identified as a significant contributor to various cardiovascular diseases (Esler, 2017), prompting individuals to monitor their stress levels in daily life. Commonly used physiological signals for stress monitoring include electrocardiography (ECG), electroencephalography (EEG), galvanic skin response (GSR), electrodermal activity (EDA), photoplethysmography (PPG), and others. PPG operates by measuring blood volume pulses through near-infrared light, providing insights into various physiological parameters (Lu et al., 2009). However, PPG has emerged as a practical alternative to ECG due to its non-invasive nature, ease of integration, ability to monitor cardiac, respiratory, and nervous system activity, cost-effectiveness, and widespread availability in commercially available wearable devices (Motin et al., 2019; 2020). The application of machine learning (ML) and deep learning (DL) techniques to analyze physiological signals has become a prominent method for stress monitoring. Several studies (Bobade & Vani, 2020; Bellante et al., 2021; Heo et al., 2021; Jahanjoo et al., 2024) have explored machine learning-based classifiers for assessing stress using either unimodal PPG signals or combinations of signals comprising PPG. However, these methods rely entirely on hand-crafted features, which can result in the omission of important signal characteristics during manual feature extraction. This limitation may hinder the generalizability and real-world applicability of such approaches.

Deep learning-based approaches are being increasingly adopted for stress monitoring, as reflected in a rising number of studies (Elzeiny & Qaraqe, 2020; Gasparini et al., 2021; Rashid et al., 2021; Kalra & Sharma, 2023). These methods excel at automatically learning relevant features directly from raw data while also allowing the integration of manually engineered features when appropriate. Most existing studies employ deep learning models that either incorporate hand-crafted features or utilize transformed signal representations such as the Fourier Transform, Short-Time Fourier Transform, or Continuous Wavelet Transform to generate input images. However, relying on hand-crafted features may lead to the loss of critical information, thereby degrading performance, while transforming signals into images significantly increases computational complexity. These limitations reduce the applicability of such methods to resource-constrained wearable devices.

In this work, we propose a dual-path hybrid CNN-BiLSTM-based shallow architecture for stress monitoring using uni-

[1]Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh. [2]Department of Information and Computer Science, SDAIA-KFUPM Joint Research Center for AI, Interdisciplinary Research Center for Bio Systems and Machines, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.. Correspondence to: Mohammod Abdul Motin <m.a.motin@ieee.org>, Mufti Mahmud <muftimahmud@gmail.com, mufti.mahmud@kfupm.edu.sa>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. 4^{th} MusIML Workshop, 2025.

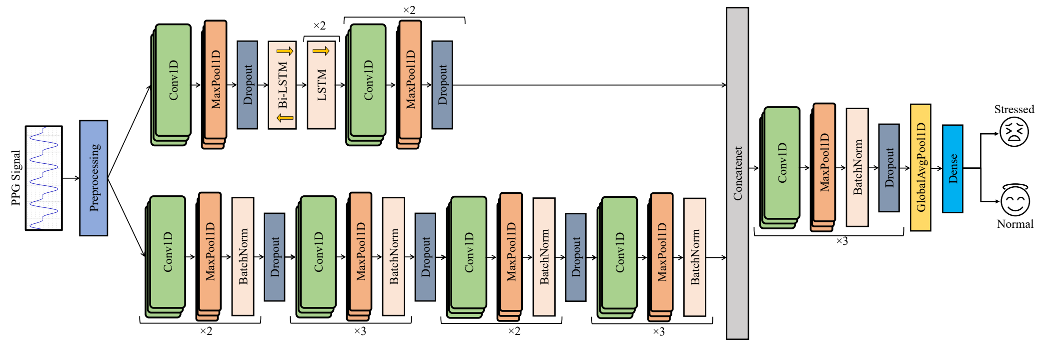

Figure 1. Architecture of the proposed hybrid CNN-BiLSTM model with dual processing paths: a convolutional-LSTM-BiLSTM pathway extracting spatio-temporal features and a convolutional pathway capturing spatial features, with concatenated outputs classified through feed-forward convolutional blocks.

modal PPG signals, specifically designed for wearable devices to achieve improved performance. To ensure interpretability of the proposed model, t-distributed Stochastic Neighbor Embedding (t-SNE) is used to visualize feature representations. The paper is structured as follows: materials and methods, results and discussion, and finally, conclusion.

2. Materials and Methods

2.1. Dataset Description

This study utilizes the publicly available Wearable Stress and Affect Detection (WESAD) dataset to evaluate the performance of the proposed model (Schmidt et al., 2018). The dataset includes physiological data, including PPG, ECG, EDA, EMG, and TEMP, collected from 17 participants who wore wearable sensors. Due to sensor malfunctions, data from two participants were excluded, leaving 15 participants (12 males and 3 females) with a mean age of 27.5 $\pm$ 2.4 years. This study exclusively focuses on PPG signals sampled at 64 Hz.

2.2. Signal Preprocessing and Augmentation

The data were segmented into 30-second non-overlapping windows, each associated with its corresponding label. To ensure consistency in signal amplitude, the PPG signals were normalized to have zero mean and unit variance. To address class imbalance within the dataset, data augmentation was applied to the minority class during the training phase. The augmentation strategy is outlined in Algorithm 1.

2.3. Deep Learning Model

The proposed dual-path hybrid convolutional neural network and bi-directional long-short term memory model, depicted in Figure 1, features a dual-path hybrid architecture designed to concurrently capture spatial and temporal dependencies in PPG signals. The first path combines convolutional, LSTM, and bidirectional LSTM layers, while the second path utilizes a series of convolutional layers. This CNN-BiLSTM framework leverages the strengths of both components: the convolutional path extracts hierarchical spatial features from the raw PPG signals, while the integrated convolutional, bidirectional, and LSTM layers model complex temporal dynamics. The model then concatenates the feature representations from both paths, preserving spatiotemporal relationships essential for accurate signal analysis. Finally, a feed-forward convolutional block performs classification based on these fused features, enabling robust pattern recognition.

In this study, the leave-one-subject-out (LOSO) cross-validation method was employed to evaluate the proposed model. The model's performance was assessed using several metrics, including accuracy, F1-score, specificity, sensitivity, AUC, and Cohen's kappa (k).

Algorithm 1 Minority Class Augmentation via Sliding Window

```
Input: Minority class dataset D_min, majority class size N_maj, sliding step size d
Output: Augmented minority class dataset D_aug
Initialize D_aug ← ∅
N_aug ← 0
while N_aug < N_maj do
    for each signal segment s in D_min do
        Apply sliding window with step size d to s
        Extract sub-segments and append to D_aug
        N_aug ← size of D_aug
        if N_aug ≥ N_maj then
            break
        end if
    end for
end while
return D_aug
```

3. Results and Discussion

The proposed model achieves an overall accuracy of 97.90%, with a specificity of 98.30%, sensitivity of 97.20%, F1-score of 97.06%, AUC of 99.12%, and Cohen's κ of 95.42%. To address class imbalance in the dataset, where normal class recordings outnumber stressed class recordings, data augmentation was applied during training. This balanced approach led to performance improvements across all evaluation metrics: accuracy increased by 2.00%, specificity by 1.60%, sensitivity by 2.70%, F1-score by 3.13%, AUC by 1.16%, and κ by 4.54%. The complete results are presented in Table 1. Figure 2 further illustrates the confusion matrix, highlighting a 5% improvement in the minority (stressed) class and an approximate 2% improvement in the majority (normal) class when augmentation is used during training. Furthermore, the Receiver Operating Characteristic (ROC) curve shown in Figure 3 demonstrates a high Area Under the Curve (AUC), indicating the model's strong discriminative capability between the two classes.

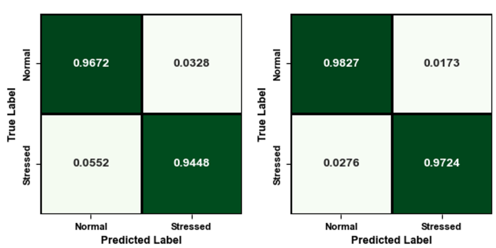

Figure 2. Normalized confusion matrices of the LOSO cross-validation without (left) and with augmentation (right).

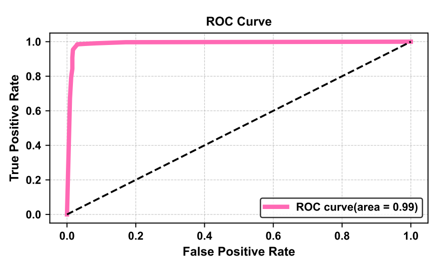

Figure 3. ROC curve illustrating an AUC as high as 0.99, indicating excellent model performance.

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 k are expressed in percentages (%).

Aug	Acc	Spe	Sen	F1	AUC	k
No	95.89	96.70	94.50	93.93	97.96	90.88
Yes	97.90	98.30	97.20	97.06	99.12	95.42

3.1. Comparison with State-of-the-Art-Architectures

A comparative analysis was conducted against both convolutional and transformer-based state-of-the-art architectures to evaluate the effectiveness of our proposed dual-path hybrid model for PPG-based stress detection. Specifically, we compared the proposed model with convolutional networks such as AlexNet, MobileNet-V1, and ResNet-18, as well as transformer-based models including CNN with Transformer (CNN+TF) and Multi-Perspective Channel Attention with Transformer (MPCA+TF) (Hu et al., 2022). As shown in Table 2, the proposed model outperforms all baseline methods across accuracy, specificity, sensitivity, and AUC on the WESAD dataset, demonstrating its superiority over the state-of-the-art-architectures.

Table 2. Performance comparison of the proposed model with state-of-the-art architectures. Accuracy, specificity, sensitivity, and AUC are expressed in percentages (%).

Model	Acc	Spe	Sen	AUC
AlexNet	94.53	96.37	91.10	95.80
MobileNet-V1	95.59	95.51	95.71	97.40
ResNet-18	96.71	95.68	98.47	98.70
CNN+TF	92.52	95.51	87.12	93.20
MPCA+TF	95.31	95.68	94.48	96.30
Proposed	**97.90**	**98.30**	**97.20**	**99.12**

3.2. Comparison with Existing Works

To benchmark the performance of the proposed model, we performed a comparative analysis with existing state-of-the-art methods for PPG-based stress classification. As shown in Table 3, our approach demonstrates superior performance compared to other methods that use only BVP signals. Although the highest accuracy reported in the literature reaches 97.2%, their study relies on multimodal signals (Bobade & Vani, 2020), our unimodal BVP-based method achieves comparable performance. The results indicate that our model either exceeds the performance of recent approaches, including those utilizing multimodal inputs.

Table 3. Comparison with the State-of-the-Art

Reference	Signal	Model	Acc
(Bobade & Vani, 2020)	PPG, ECG, EDA, EMG, RESP	ANN	95.21%
(Bellante et al., 2021)	PPG, EDA, RESP	ANN	97.20%
(Heo et al., 2021)	PPG	LDA	96.50%
(Rashid et al., 2021)	PPG	DNN	88.56%
(Jahanjoo et al., 2024)	PPG	SVM	95.55%
Proposed	**PPG**	**DL**	**97.90**%

3.3. Model Interpretability

Deep learning models typically function as black boxes, making visualization of their internal feature representation essential for physiological data analysis. The feature learning process of the dual path hybrid CNN-BiLSTM model is visualized using t-SNE as illustrated in Figure 4. The input layer shows complete overlap between normal (red) and stressed (blue) classes, confirming the inherent inseparability. Through successive transformations, the model develops discriminative features, with the global average pooling layer achieving clear class separation. This progression demonstrates our architecture's ability to extract meaningful patterns from initially indistinguishable inputs.

3.4. Compatibility with Wearable Devices

Continuous stress monitoring is crucial for maintaining physical and mental well-being (Esler, 2017). Photoplethysmography (PPG) sensors, being cost-effective and widely integrated in commercial wearables (Lu et al., 2009), offer a practical solution for daily stress assessment. However, wearable devices impose strict constraints on memory usage and computational capacity, necessitating efficient model architectures. Our dual path hybrid CNN-BiLSTM model addresses these requirements with a compact parameter count of 1,386,306 (5.29 MB), making it suitable for deployment on resource-constrained wearable platforms.

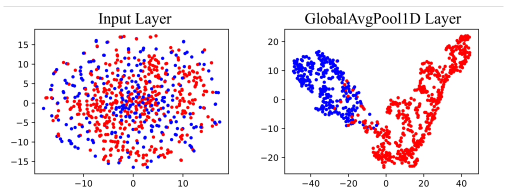

Figure 4. The t-SNE feature map visualization illustrates the feature learning process of the hybrid CNN-BiLSTM model. Initially, the features were indistinguishable; however, as training progressed, a clear separation emerged between the normal (red) and stressed (blue) classes, demonstrating the model's effectiveness in learning discriminative representations.

4. Conclusions

In this study, we propose a lightweight hybrid CNN-BiLSTM architecture for stress detection using solely PPG signals from wearable devices. The proposed model achieves exceptional performance with 97.90% accuracy, 98.30% specificity, 97.20% sensitivity, 97.06% F1-score, 99.12% AUC, and 95.42% Cohen's k, demonstrating superiority over existing approaches in the literature. With minimal parameters, the architecture is highly efficient and well-suited for resource-constrained wearable devices. These results highlight the potential for accurate, continuous, and non-invasive stress monitoring in practical applications. Future work will explore multi-class stress classification and real-world system integration.

References

Bellante, A. et al. Emocy: Towards physiological signals-based stress detection. In *Proc. BHI*, pp. 1–4, 2021.

Bobade, P. and Vani, M. Stress detection with machine learning and deep learning using multimodal physiological data. In *Proc. ICIRCA*, 2020.

Elzeiny, S. and Qaraqe, M. Stress classification using photoplethysmogram-based spatial and frequency domain images. *Sensors*, 20(18):5312, 2020.

Esler, M. Mental stress and human cardiovascular disease. *Neurosci. Biobehav. Rev.*, 74:269–276, 2017.

Gasparini, F., Grossi, A., and Bandini, S. A deep learning approach to recognize cognitive load using ppg signals. In *Proc. PETRA*, pp. 489–495, 2021.

Heo, S., Kwon, S., and Lee, J. Stress detection with single ppg sensor by orchestrating multiple denoising and peak-detecting methods. *IEEE Access*, 9:47777–47785, 2021.

Hu, S., Cai, W., Gao, T., and Wang, M. A hybrid transformer model for obstructive sleep apnea detection based on self-attention mechanism using single-lead ecg. *IEEE Transactions on Instrumentation and Measurement*, 71: 1–11, 2022.

Jahanjoo, A., TaheriNejad, N., and Aminifar, A. High-accuracy stress detection using wrist-worn ppg sensors. In *Proc. ISCAS*, pp. 1–5, 2024.

Kalra, P. and Sharma, V. Mental stress assessment using ppg signal a deep neural network approach. *IETE J. Research*, 69(2):879–885, 2023.

Lu, G., Yang, F., Taylor, J., and Stein, J. A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects. *J. Med. Eng. Technol.*, 33(8):634–641, 2009.

Motin, M. A., Karmakar, C. K., and Palaniswami, M. Selection of empirical mode decomposition techniques for extracting breathing rate from ppg. *IEEE Signal Process. Lett.*, 26(4):592–596, 2019.

Motin, M. A., Karmakar, C., Palaniswami, M., and Penzel, T. Photoplethysmographic-based automated sleep–wake classification using a support vector machine. *Physiol. Meas.*, 41(7):075013, 2020.

Rashid, N. et al. Feature augmented hybrid cnn for stress recognition using wrist-based photoplethysmography sensor. In *Proc. EMBC*, 2021.

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proc. ICMI*, pp. 400–408, 2018.