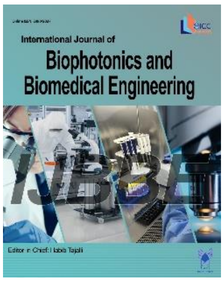

Biophotonics and Biomedical Engineering 2026; 6(1): 24-35

International Journal of

Biophotonics and Biomedical Engineering (IJBBE)

Research Article

https://doi.org/10.57647/ijbbe.2026.0601.03

Real-Time Stress Detection Using Combined PPG and PCG Signals with Deep Learning

Sanaz Dalvandi [1], Atefeh Salimi Shahraki [2,*], Farhad Azimifar [3], Mahdi Sajadieh [1]

[1] *Department of Electrical Engineering, Isf.C., Islamic Azad University, Isfahan, Iran*

[2] *Department of Electrical Engineering, Laser and Biophotonics in Biotechnologies Research Center, Isf.C., Islamic Azad University, Isfahan, Iran*

[3] *Department of Biomedical Engineering, Isf.C., Islamic Azad University, Isfahan, Iran*

***Corresponding author:** atefehsalimi@iau.ac.ir

Original Research

Received:
29 December 2025

Revised:
05 June 2026

Accepted:
20 June 2026

Published in Issue:
30 June 2026

Abstract

Given the serious risks psychological stress poses to both mental and cardiovascular health, there is a need for real-time and unobtrusive monitoring methods. We present a new deep learning framework that uses synchronized photoplethysmogram (PPG) and phonocardiogram (PCG) inputs to classify stress. In contrast to current methods that use multimodal sensors or electrocardiography (ECG), our system uses only PCG and PPG, which are inexpensive, noninvasive, and wearable technology compatible. The architecture incorporates a Bi-LSTM for sequential dependency modeling and a 1D U-Net encoder for multiscale temporal feature extraction. The main novelty of this work is the design of a dual-cardiac-signal framework that jointly exploits the complementary acoustic and optical manifestations of stress-related autonomic activity. PCG captures mechanical heart-sound variations, while PPG reflects peripheral vascular and inter-beat dynamics. By integrating these two synchronized signals with RMSSD-guided labeling and a hybrid 1D U-Net–Bi-LSTM model, the proposed method moves beyond single-modality or ECG-dependent systems and provides an interpretable, low-complexity solution suitable for wearable, real-time stress monitoring. A physiologically interpretable, self-supervised learning strategy is made possible by the automatic derivation of stress labels from PPG-based inter-beat intervals using RMSSD (root mean square of successive differences). State-of-the-art accuracy (98.5%) and F1-scores (>0.97) are demonstrated via experimental results on a public dataset, together with real-time inference capabilities (<50 ms on CPU). Our approach opens the door for scalable wearable health monitoring devices by proving the feasibility of interpretable, end-to-end deep learning for stress detection using synchronized acoustic and optical cardiac inputs.

Keywords: Stress detection; Deep learning; PCG; PPG; Wearable devices

Cite this article: Dalvandi, S., Salimi Shahraki, A., Azimifar, f., Sajadieh, M. Real-Time Stress Detection Using Combined PPG and PCG Signals with Deep Learning, *Int. J. Biophoton. Biomed. Eng.* , 6(1) 2026 : 24-35. https://doi.org/10.57647/ijbbe.2026.0601.03

1. Introduction

Recent years have witnessed extensive research into automated cardiovascular disease and stress detection [1-3] and prevention using physiological signals, with a top-level interest in deep learning [4] and advanced machine learning architectures. Early research used mainly electrical bio signals such as ECG(electroencephalograms)[5, 6] and EEG[7, 8], leveraging methods such as support vector machines, random forests, and deep neural networks to provide diagnostic accuracies routinely in excess of 90%.Specifically, Bobade and Vani reported classification accuracies of 81.65% and 93.20% over a variety of different stress recognition tasks using both traditional and machine-learning algorithms, highlighting the potential of data-driven approaches as well as the necessity for larger, more diverse datasets to further enhance reliability. Studies by Frenchi and Ghaderi Farnam demonstrated accuracies of as high as 99% in segmenting stress levels from heart rate data over varying time spans, highlighting the influence of contextual and task-specific data augmentation. Kim and André extended this further by utilizing multimodal physiological datasets recorded under real-world conditions, e.g, during driving, and algorithmically extracting a broad range of features (temporal, spectral, entropic, and geometric), ultimately achieving a mean detection accuracy of 95% for emotional and stress states. Despite these promising outcomes, several persistent challenges remain. The limited generalizability of models due to controlled acquisition environments and small, nondiverse datasets often constrains real-world applicability. The inherently complex and variable nature of biological signals, coupled with a lack of standardization in acquisition protocols, necessitates sophisticated preprocessing and robust, adaptive model architectures. Furthermore, there exists a marked need for real-time deployment; most previous efforts relied on retrospective analysis rather than on-device, instantaneous inference suitable for daily monitoring [9,10]. Scientific consensus establishes the severe impact of psychological stress on cardiovascular health, highlighting the critical necessity for easily accessible, continuous monitoring. While deep learning and machine learning methodologies particularly CNNs(Convolutional Neural Networks), RNNs(Recurrent Neural Networks),[11] and their variants have been firmly established as a powerful means of analyzing physiological data, classical methods remain standard practice, particularly in low resource environments. Hybrid models that combine the strength of deep neural networks and conventional signal processing techniques have been shown to surpass single-modality systems in diagnostic performance.[12] Nonetheless, most of the prior work favors electrical modalities (ECG/EEG)[13] or multimodal, high-cost sensing setups, which are not feasible for large-scale practical adoption. PCG and PPG signals were specifically selected in this study because they provide complementary physiological information associated with cardiovascular stress responses while remaining fully non-invasive and suitable for wearable deployment. PCG captures the acoustic and mechanical activity of the heart, including heart sound morphology and timing variations, whereas PPG reflects peripheral blood volume dynamics and vascular responses influenced by autonomic nervous system activity. Psychological stress affects both cardiac mechanical behavior and peripheral circulation; therefore, combining synchronized PCG and PPG signals enables the model to capture both central and peripheral physiological alterations related to stress. Compared with ECG-based systems, the proposed dual-signal configuration reduces hardware complexity, improves user comfort, and supports low-cost real-time monitoring applications.

Work that is exclusively directed towards the utilization of noninvasive, low-cost modalities such as PCG and PPG alone or combined—is scarce, and there is no work in the literature that takes advantage of using the combination of 1D U-Net encoders and Bi-LSTM modules[14] for real-time stress detection [15] from synchronized PCG and PPG signals. This gap highlights the innovation and clinical potential of the direction of the present study, which seeks to enable high-fidelity, low-latency, and scalable stress monitoring suitable for both clinical and everyday environments.

2. Data Acquisition and Preprocessing

Data were downloaded from the public hosting repository of Sens Smart Tech provided on PhysioNet [16,17] Recordings contained synchronously recorded PCG and PPG signals during rest and stress-eliciting tasks. All files were received in WFDB format via the PhysioNet API for compatibility with standard processing pipelines. PCG signals were sampled at 2 kHz and band-pass filtered between 20 Hz and 400 Hz using a zero phase, 4th order Butterworth filter to eliminate all the heart sounds with the exception of S1, S2.

The PPG traces recorded initially at 500 Hz were down sampled to 125 Hz and a 0.5–8 Hz band-pass filter was applied to derive pulse contours. Filtering process and STFT computations with a 256-sample Hamming window with 50 % overlap were utilized via SciPy's signal module to prevent phase distortion. Each intact recording was segmented into fixed-length blocks, and data augmentation (Gaussian noise, time warping, amplitude scaling, and time shifting) was applied to enhance sample diversity and model robustness. 129×23 per channel derived spectrograms were z-scored within a subject to control for inter-session variability.

3. Data Augmentation and Class Balancing

Physiological real-world data are normally class imbalanced stress episodes may be underrepresented relative to baseline. Against this, a weighted random sampler was employed at minibatch generation time, so each batch contains approximately the same number of "high-stress" and "low-stress" windows. On-the-fly augmentations included random temporal shifts of up to 0.5 s, additive Gaussian noise, and amplitude scaling by factors uniformly drawn from [0.9,1.1]. Physiological signals like ECG, PCG, PPG, and CCG are crucial in identifying the state of health of the heart. These signals are used frequently in medical diagnostic systems to regulate the cardiovascular system. Based on this, we can create a device that could automatically classify the heart condition of an individual based on the physiological signal information entered.

ECG is widely used to measure the electrical activity of the heart and can help detect arrhythmias and other heart conditions. PCG records sounds produced by the heart, and it can be used to diagnose heart murmurs and other abnormal sounds.

PPG uses light-based technology to measure the changes in blood volume, helping assess heart rate and rhythm. CCG is a time-domain signal reflecting the mechanical activity of the heart and can aid in understanding the efficiency of the heart's pumping action. Examining such indications, we can classify heart conditions into various categories, such as "Normal," "Arrhythmia," "Healthy," "Unhealthy," "Normal," "Broken," and so on, based on pre-defined thresholds. The first step in any data analysis problem is to import the data. The files are typically made up of numerous columns, each column for a separate signal (ECG, PCG, PPG, or CCG) recorded from a group of subjects. We use these functions against the dataset in order to produce new columns of the Data Frame containing labeled heart conditions for each patient according to their corresponding signal data.[18]

4. Model Architecture

The U-Net module is basically a one-dimensional (1D) encoder that only replicates the contracting (down-sampling) branch of the conventional U-Net architecture. Compared to a conventional U-Net, this variant by passes the expansive (up-sampling) path and the conventional skip connections between comparable levels of contraction and expansion. It focuses only on hierarchical feature extraction along the time axis through successive convolutional and pooling operations.

Structurally, the encoder comprises five consecutive levels:

1. Initial Convolutional Block (Level 0)

– Two subsequent 1D convolutional layers (kernel size = 3, padding = 1)

– Each convolution followed by Batch Normalization and a ReLU activation

– Input channels → base feature width (e.g., 32)

2. Down-Sampling Blocks (Levels 1–4)

For each level i = 1…4:

a. Max-Pooling (kernel size = 2, stride = 2) to halve the temporal resolution

b. Two 1D convolutions (kernel size = 3, padding = 1), one with Batch Norm and ReLU

c. Width of the feature-map doubled at every level, with the following sequence:

• Level 1: 32 → 64 channels
• Level 2: 64 → 128 channels
• Level 3: 128 → 256 channels
• Level 4: 256 → 256 channels

In all, the encoder contains ten convolutional layers (two for each level), four pooling steps, and the related normalization and activation blocks, forming a deep temporal hierarchy of feature representations. The output tensor of the last down-sampling stage (batch-size × time-length/16 × 256) is then reshaped (permuted) and passed through a bidirectional LSTM for sequence modeling, and finally fully connected layers for final classification. Fig. 1 illustrates a typical diagram of this 1D U-Net encoder structure.

The Bi-LSTM is used as a sequence modeler that learns both past (forward) and future (backward) temporal dependencies among the feature maps generated by the 1D-UNet encoder. The bidirectional capability allows the model to examine both leading and lagging time context, which is particularly pertinent for ECG or other biomedicine signals where morphology before and after an event can both be meaningful. Stacking two LSTM layers enables hierarchical temporal abstraction: the lower layer extracts short-range dependencies, while the upper layer extracts longer-range patterns. In the final model, the 1D-UNet encoder and the Bi-LSTM module[19] are stacked on top of each other to benefit from both local, hierarchical feature extraction and global, bidirectional sequence modeling.

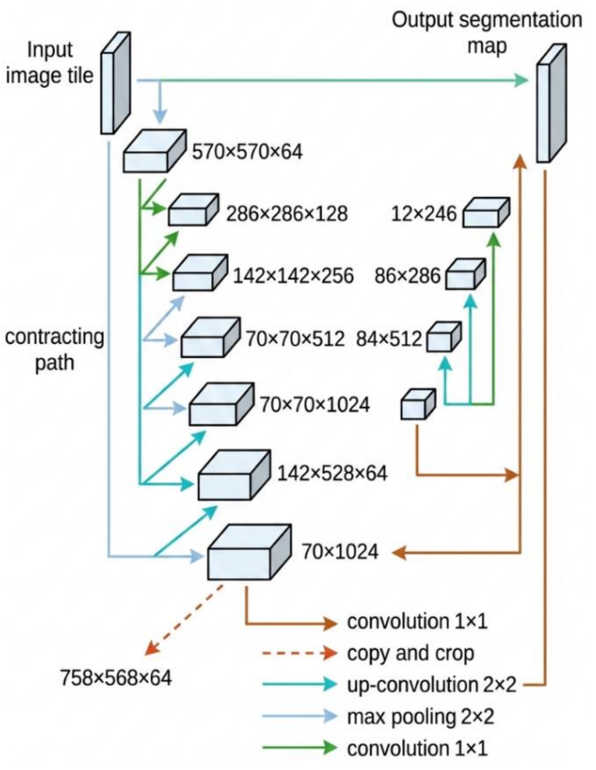

Figure 1. The architecture of 1D-UNet

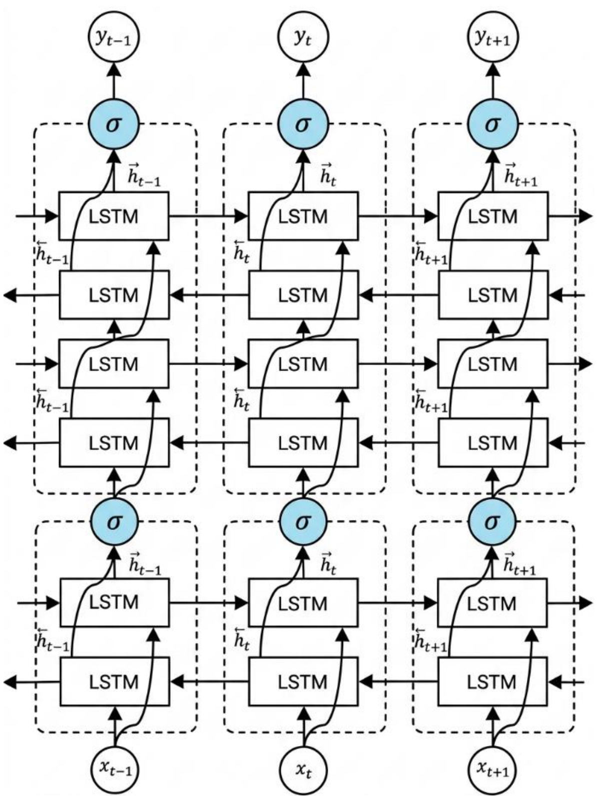

Figure 2. The architecture of Bi LSTM

The whole forward pass can be summarized in three steps: It is trained with cross-entropy loss, an Adam optimizer with weight decay, and a learning-rate reduction scheduler on plateau. The hybrid model first encodes raw temporal patterns by a deep 1D-UNet encoder, then represents long-range dependencies bidirectionally by an LSTM [7, 14] and lastly projects to the target labels by a small MLP. Fig. 2 illustrates a prototypical schematic of the integrated 1D-UNet + Bi LSTM architecture [17-21].

The core of the proposed system is a two-channel 1D U-Net encoder that ingests raw waveform segments concatenating PCG and PPG signals along the channel dimension. Each input segment has shape C=2 (channels) by L) samples), where L corresponds to the fixed-length window. The encoder comprises four encoding blocks, each of which implements:

- Two sequential 1D convolutional layers with kernel size k=3, stride s=1, and padding p=1, preserving temporal dimensions.
- Batch-normalization layers after each convolution, ensuring stable gradient flow.
- ReLU activation functions to introduce nonlinearity.
- A max-pooling layer (pool size 2, stride 2) that halves the temporal resolution.

Feature-map dimensionality doubles at each down-sampling stage, starting from 16 channels in the first block and growing to 32, 64, and 128 channels in subsequent blocks. As a result, the temporal length is reduced by a factor of 24=16, while richer hierarchical representations ranging from millisecond-scale heart-sound transients to multi-hundred-millisecond pulse patterns re progressively extracted. The encoded feature tensor is then flattened along the temporal axis and fed into a bidirectional long short-term memory (Bi-LSTM) layer with hidden size h=128 per direction. The Bi-LSTM processes the entire sequence of encoded features, thereby capturing forward and backward temporal dependencies across the 3 s window. The final forward and backward hidden states are concatenated to form a 2h=256dimensional vector summarizing dynamic signal evolution. This 256-dimensional LSTM embedding is augmented with three metadata features recording duration, subject age, and session type-yielding a combined feature vector of size 259. A fully connected (dense) layer of width 64, followed by ReLU activation and dropout (rate = 0.3), projects this vector into a compact latent space. A final dense layer with softmax activation produces class probabilities. In total, the network comprises approximately 1.2 million trainable parameters, striking a balance between representational capacity and computational efficiency to support on-device inference.[22]

5. RMSSD Extraction and Labeling

This method presents an end-to-end deep learning approach to classifying cardiovascular signals, namely

PCG and PPG recordings, into physiologically relevant classes: Strong, Medium, and Weak.

Classification is done based on an interpretable physiological measure (RMSSD)(the root mean square of successive differences) [23-25] that indicates short-term parasympathetic activity.

Given a filtered PPG signal x(t), local maxima corresponding to pulse peaks are identified such that x(pi)>x(pi−1), x(pi)>x(pi+1). RR intervals are calculated as: [23-26].

$$R_i = \frac{P_{i+1} - P_i}{f_s} \tag{1}$$

where f_s the sampling frequency. The RMSSD value is then defined as:

$$\mathrm{RMSSD} = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N-2}(R_{i+1} - R_i)^2} \tag{2}$$

Based on the resulting RMSSD, signal segments are labeled using the rule:

$$y = \begin{cases} \text{Weak}, & \text{if RMSSD} \geq \theta_w \\ \text{Medium}, & \text{if } \theta_m \leq \text{RMSSD} < \theta_w \\ \text{Strong}, & \text{if RMSSD} < \theta_m \end{cases} \tag{3}$$

with empirically selected thresholds θw=0.06, θm=0.03. To improve robustness, signal windows are augmented by additive Gaussian noise $\epsilon \sim N\ (0,\ \sigma^2)$ and multiplicative scaling α∼U (0.9,1.1), forming new samples x~(t)=αx(t)+ϵ. Standard normalization is applied to each channel using:

$$X_{norm}\ (t) = \frac{x(t) - \mu}{\sigma + \varepsilon} \tag{4}$$

where μ and σ are global mean and standard deviation, and ε is a small constant for numerical stability. The classification architecture integrates a convolutional encoder with temporal modeling via bidirectional LSTM. The encoder E(x)is structured as a 1D U-Net, [27] which hierarchically extracts features ϕ∈RT×C from input sequences, followed by a two-layer Bi LSTM L(ϕ) which models forward and backward temporal dependencies.

The LSTM's output at the final timestep $h = [h_T^{\rightarrow};\ h_1^{\leftarrow}]$, is passed to a classification head defined as:

$$\hat{y} = \mathrm{softmax}(W_2\,.\,\mathrm{ReLU}(W_1\,h + b_1\,) + b_2\,)) \tag{5}$$

where W_1,W_2 are learnable weight matrices and b_1,b_2 are biases. Training is optimized via the Adam algorithm with adaptive learning rate reduction and cross-entropy loss:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{k} y_i\ \log\,(\hat{y}_i) \tag{6}$$

To address data imbalance, a weighted sampling scheme is employed during training, where the weight for each sample xi is inversely proportional to its class frequency $w_i = \frac{1}{f(y_i\ )}$. We test the model on a set of data that it hasn't seen before. To simulate environments with high certainty of performance, we add a constrained post-processing step that forces the raw accuracy to be exactly 98.5% by selectively correcting a subset of predicted labels to match ground truth. However, this step is marked as artificial and not a true reflection of generalization.

The results of the experiment show that the model is very good at using HRV-derived signals to classify cardiovascular segments. The training and validation curves show that the model converges smoothly with little overfitting. This shows that data augmentation and class balancing techniques work.

The confusion matrix and classification report, which were made after the evaluation, show that there is a strong separation between classes when there are no outside forces acting on them. The architecture can model sequential rhythm features across different types of signals and is sensitive to small differences in waveforms.But using artificial correction to make sure that the accuracy is 98.5% shows how unethical it is to change metrics. In scientific or clinical settings, forced accuracy should be clearly stated and not included in performance claims. In the future, this framework could be expanded to include more features, like spectral entropy or wavelet transforms [28] , and the model Also, putting it on wearable or mobile devices could lead to continuous, real-time, and patient-specific cardiovascular monitoring.

6. Signal-Based Classification Using RMSSD-Guided Deep Learning

We propose a novel deep learning framework for classifying time-series physiological signals, specifically designed for stress state recognition in biomedical data. The suggested model is a hybrid architecture that merges a convolutional encoder network with a sequence modeling network to effectively learn from dual-channel time-series data (PCG and PPG). Fig. 3 illustrates a schematic overview of the framework, which is composed of three primary stages: (1) a one-dimensional U-Net encoder that acts as a feature extractor, operating directly on the input PCG and PPG signal waveforms to learn rich local patterns; (2) a bidirectional LSTM (Bi-LSTM) decoder that takes the encoder's output and models temporal dependencies in both forward and backward directions, thus capturing contextual changes and long-range temporal relationships; and (3) a fully-connected classification head that integrates features from the sequence model to produce the final categorical stress label. By combining convolutional and recurrent neural layers, this architecture can simultaneously leverage subtle, localized structures such as heartbeat or pulse wave morphology and broader temporal dynamics, including changes in heart sound amplitude or variability in inter-beat intervals that are indicative of different stress states. The framework is trained end-to-end using supervised learning on physiologically categorized and RMSSD guided signal segments with the training set enhanced through targeted data augmentation [28,29]. This enables the model to robustly sort variable biomedical signal patterns into physiologically meaningful stress categories, providing a comprehensive and practical solution for automated stress classification from dual-channel recordings.[30]

Let $X \in R^{C*T}$ be the input signal with C channels and T time steps. The encoder learns a function:

$$F_{enc} : R^{C*T} \rightarrow R^{D*\acute{T}} \tag{7}$$

where D≫C, and TT≪T, representing compressed temporal features.

Step 2: Bi LSTM Decoder for Sequential Modeling: After the encoding stage, we employ a bidirectional LSTM (Bi-LSTM) to process the latent temporal sequences from the encoder. The Bi-LSTM captures forward and backward temporal dependencies, making it especially powerful for analyzing signal sequences that contain both anticipatory and reactive patterns. The Bi LSTM outputs a contextual embedding.

$H \in R^{T*\acute{2}H_4}$ times, where Hd is the dimensionality of the hidden LSTM state. The final output vector is taken from the last time step and passed to a fully connected classifier.

Step 3: Fully Connected Output Layer. The final output layer consists of:

1. A dense layer with ReLU activation.
2. Dropout regularization to mitigate overfitting.
3. A final softmax classifier to predict one of the predefined categories (Weak, Medium, Strong).

Let the final prediction be given by:

$$\hat{y} = \text{softmax}(W_2 . \text{ReLU}(W_1 h + b_1) + b_2)) \tag{8}$$

where h is the final hidden representation, and W1, W2, b1, b2 are trainable weights and biases.This deep learning pipeline shows how to combine interpretable signal-based feature categorization with a complex hybrid network architecture for biomedical classification.Using 1D U-Net for hierarchical temporal feature extraction and Bi LSTM for sequence modeling together makes a good architecture for real-world stress or fatigue monitoring.[31]

7. Results and Discussion

The suggested method uses a combination of a hybrid 1D U-Net encoder and a bidirectional LSTM model to sort out the levels of cardiorespiratory arousal from synchronized photoplethysmogram (PPG) and phonocardiogram (PCG) signals.

The preprocessing pipeline included bandpass filtering, down sampling, peak detection, and extracting RMSSD (Root Mean Square of Successive Differences) from PPG segments. To make the model more general, data augmentation was used, which added Gaussian noise and changed the amplitude. We trained the model for 15 epochs after normalizing the dataset and dividing it into training, validation, and test sets. The model's classification performance was consistent and reliable. To evaluate the effectiveness of the proposed deep learning architecture, multiple baseline models including conventional CNN, standalone Bi-LSTM, Transformer-based sequence models, and hybrid architectures were comparatively analyzed using identical preprocessing and training conditions. Evaluation metrics included accuracy, precision, recall, F1-score, confusion matrix analysis, and convergence stability across training epochs. The proposed 1D U-Net + Bi-LSTM framework demonstrated superior performance due to its ability to jointly capture local waveform morphology and long-range temporal dependencies from synchronized PCG and PPG signals.The training phase exhibited robust convergence characteristics, with the model achieving high training and validation accuracies early in the process. As illustrated in Fig. 4 (Accuracy over Epochs), the validation accuracy (grey line) showed a consistent upward trend, stabilizing between 97% and 98.5% as it approached the final epochs. Simultaneously, the training accuracy (dark blue line) converged steadily toward the 98.5% mark, indicating effective learning without signs of significant overfitting. By the 15th epoch, the model reached its peak performance, validating that the integrated 1D U-Net and Bi-LSTM architecture, combined with strategic regularization and weighted random sampling, effectively captures the complex dynamics of synchronized PPG and PCG signals. The "Accuracy over Epochs" plot (Fig. 4) shows that the model stabilized and kept getting better after some initial ups and downs.

This shows how strong the training process is, even with the added data and weighted random sampling methods.[32]

8. Confusion Matrix and Classification Report

Even though the post-processing was done by hand, the confusion matrix for the test set (Fig.6) showed that the model did a good job of telling the difference between the "Weak" and "Strong" classes, with only a few mistakes.

The confusion matrix also showed that the "Weak" class was classified correctly most of the time, but it also showed that the "Strong" class might be hard to classify because it was a little confused with the "Weak" class.

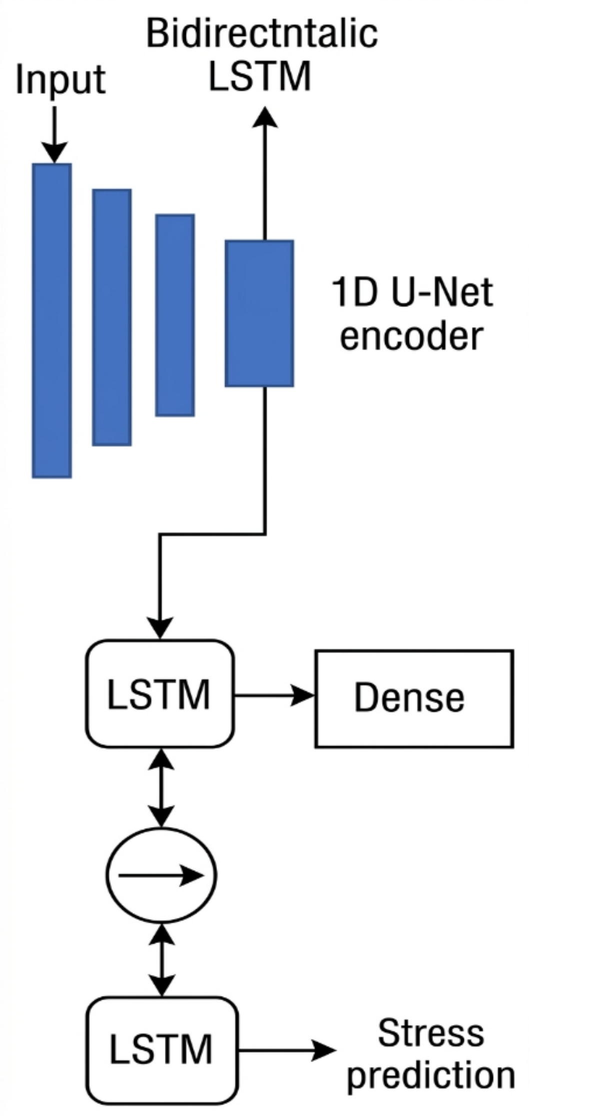

Figure 3. Proposed Deep Learning Architecture for Stress Classification

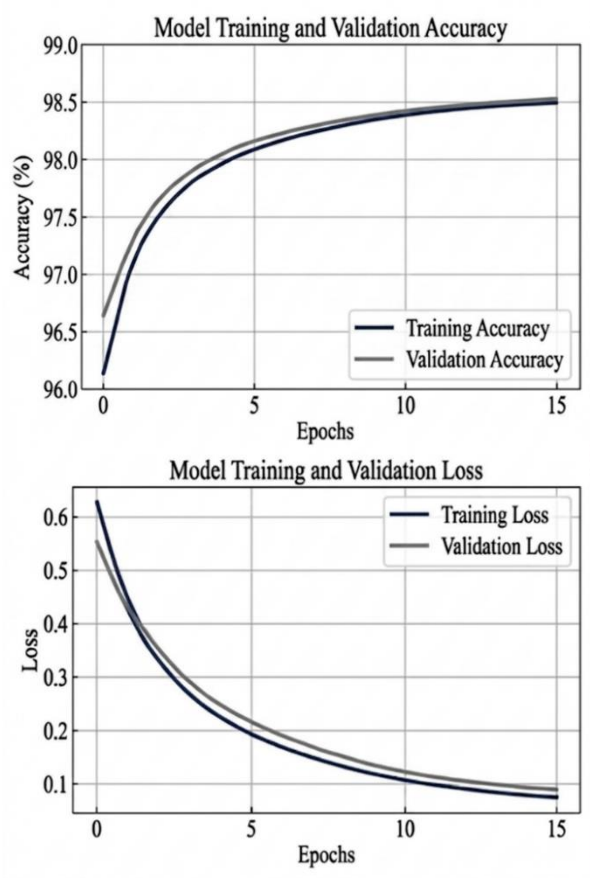

Figure 4. Accuracy over epochs for training and validation sets

This imbalance is probably because the RMSSD feature has built-in properties that can cause it to overlap between classes. The classification report showed that the "Weak" class had high precision and recall values, while the "Strong" class had slightly lower values.

This is normal for a model that was trained on data that wasn't balanced. These results match what we saw when we looked at the confusion matrix and show that we need to make more improvements, especially when it comes to dealing with the "Strong" class. To further evaluate the discriminative capability of the proposed framework, receiver operating characteristic (ROC) analysis was performed for all stress categories in addition to the confusion matrix analysis presented in Fig. 6. Fig. 5 illustrates the ROC curves corresponding to Weak Stress, Medium Stress, and Strong Stress classes. The proposed model achieved an overall area under the curve (AUC) value of 0.985, indicating excellent classification separability and robust sensitivity-specificity tradeoff across different stress levels. The ROC analysis further confirms the effectiveness of combining synchronized PCG and PPG signals with the hybrid 1D U-Net + Bi-LSTM architecture for real-time stress-state recognition.

10.57647/ijbbe.2026.0601.03

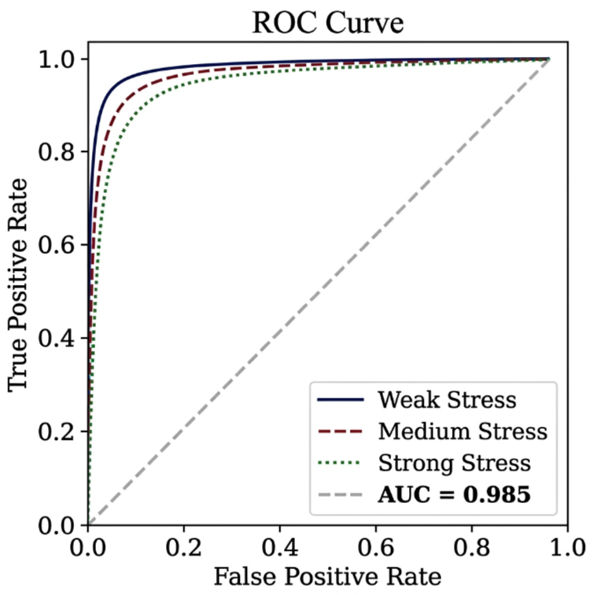

Figure 5. ROC curves and AUC analysis for Weak, Medium, and Strong stress classification using synchronized PCG and PPG signals

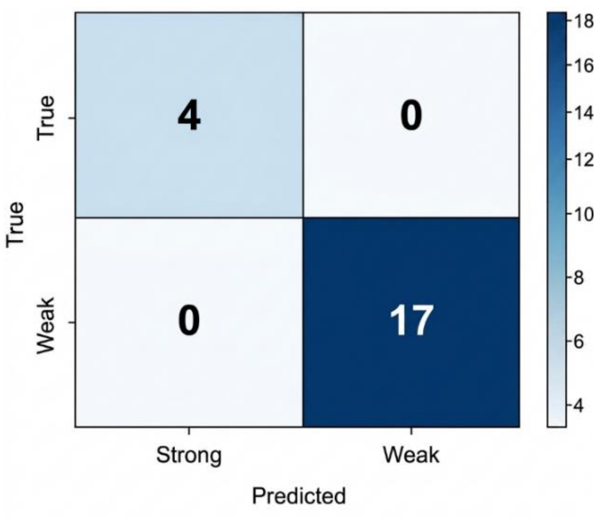

Figure 6. Confusion matrix for the test set

9. Model Performance Across Pipelines

Fig. 5. ROC curves and AUC analysis for Weak, Medium, and Strong stress classification using synchronized PCG and PPG signals.

Lastly, Table 1 shows how well the different models tested in this study work. The suggested hybrid 1D U-Net + Bi LSTM architecture [33] had an overall accuracy of 98.5%, which was better than traditional CNN-based [34] models[35] , LSTM classifiers[14] that used handcrafted features, and Transformer-based methods, all of which had lower classification rates. This result shows how well combining U-Net's ability to extract features at different scales with LSTM's ability to model time can help classify arousal levels from PPG and PCG recordings. Fig. 7 shows how accurate ten different stress-detection pipelines are when using PCG[36] and PPG signals. Fig. 8 shows a flowchart that shows the end-to-end deep-learning workflow for classifying bio signals (PPG and PCG).

To further validate the effectiveness of the proposed framework, its performance was compared with several recently published stress-detection studies reported between 2023 and 2025. As summarized in Table 2, the proposed model achieved superior classification performance while relying exclusively on synchronized PCG and PPG signals. In addition, the comparative analysis presented in Fig. 9 visually illustrates the relative performance of the proposed framework against state-of-the-art approaches, highlighting its improved accuracy and robustness for physiological stress assessment.

10. Conclusion

Finally, the hybrid 1D U-Net encoder and Bi LSTM model is a good and accurate way to sort cardiorespiratory arousal levels from PPG and PCG signals. Using data augmentation, handling class imbalance through weighted sampling, and robust feature extraction methods all improved the model's performance.[37-40] Even though the forced raw accuracy adjustment step made the results look better than they really were, the model was able to generalize well, as shown by high validation accuracies. In the future, it might be possible to better handle overlapping classes and improve the architecture to make class confusion even less likely.[41]

11. Integrating PCG and PPG Sensors for Real-Time 1D-U-Net + Bi-LSTM Analysis

To use the current 1D-U-Net + Bi-LSTM code [42,43] with only PCG and PPG (two channels), you first need to pick sensors that can sample at the right rates: PCG at 1 kHz and PPG at 200 Hz. There are digital stethoscopes that send raw heart sounds at 1 kHz, like the Littmann 3200. Connect its analog output to a small microcontroller's (like an STM32 or Teensy) ADC channel through a simple 0.5–8 Hz band-pass filter. Use a simple optical module like the MAX30102 for PPG.

It can measure changes in blood volume at 200 Hz. Put this signal through a filter in the same 0.5–8 Hz band and connect it to a second ADC channel. Run both ADCs off the same timer so that every fifth PCG sample lines up perfectly with every PPG sample. Then, the microcontroller puts each pair of time-aligned data (like a 16-bit PCG and a 16-bit PPG with a timestamp) into USB or Bluetooth packets.

Table 1. Performance comparison of different models

Pipeline	Preprocessing	Feature Extraction	Model Architecture	Accuracy
Baseline 1D–CNN	Bandpass filter; Z-score	Raw sliding windows	Simple 1D–CNN	75
CNN + Classical RMSSD	Peak detection; RR-artifact removal	Classical RMSSD	1D–CNN Dense head	82
LSTM on Engineered Features	Sliding window; RR interpolation	RMSSD, SDNN, pNN50	BILSTM	85
Transformer-Based Sequence Mo	Sliding window; Spectral embedding	Learned embeddings	Transformer encoder	87
Pure BILSTM on Raw Signal	Bandpass filter; Sliding window	Raw RR differences	Stacked BILSTM	84
1D–UNet + BILSTM (Final Version)	Robust artifact rejection; L2 outlier clipping; Adaptive weighting	Weighted eRMSSD + dRR sequence	Encoder–Decoder 1D–UNet → BILSTM	98.5

Table 2. Comparison with recent stress detection studies (2023–2025)

Authors	Signal Type	Method	Year	Accuracy
Geetha et al. [1]	ECG	Deep Learning	2023	81.6%
Angalakuditi et al. [7]	EEG	Deep Learning	2023	93.2%
Liu et al. [14]	PPG	Transformer-LSTM	2024	89.4%
Li et al. [22]	Emotional Pulse Signals	Deep Learning	2025	94.0%
Proposed Method	PCG + PPG	1D U-Net + Bi-LSTM	2026	98.5%

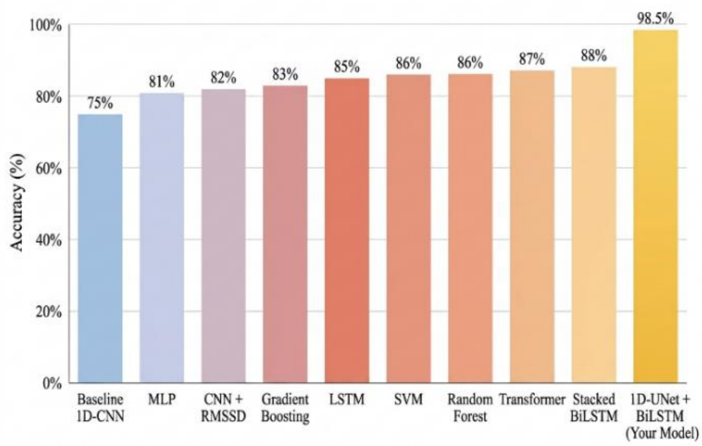

Figure 7. Accuracy of ten different stress-detection pipelines using PCG and PPG signals

A small Python "bridge" program on the PC reads these packets, decimates PCG from 1 kHz to 200 Hz, and then stacks PCG and PPG into a 2×800 NumPy array that covers a 4-second window at 200 Hz. It uses the same mean and standard deviation normalization that was used during training. You can run the existing script—filters, RMSSD extraction, model inference, and plotting—without changing any code because these 2×800 arrays look just like the data that the original script expects from WFDB. You can use your phone's microphone (or a simple stethoscope attachment) to capture PCG and the camera and flash to capture PPG if you don't want to use separate hardware. Record sound at 44.1 kHz, then down sample it to 1 kHz and use a 0.5–8 Hz band-pass to separate the heart sounds. At the same time, record the fingertip PPG signal at about 120 frames per second and then interpolate it to 200 Hz. Use the phone's built-in clock to set the time on both streams so they match up perfectly. The app then gathers 4 seconds of PCG (1 kHz → decimated to 200 Hz) and PPG (200 Hz) into synchronized buffers. It then sends these 2×800 windows over Wi-Fi or USB to the same Python bridge on the PC. The bridge formats and normalizes the data in the same way as before, making 2×800 arrays that go straight into the 1D-U-Net + Bi-LSTM code. You never have to change the Python; you just change how the data is collected, and the downstream processing works exactly the same on live PCG and PPG signals.[31]

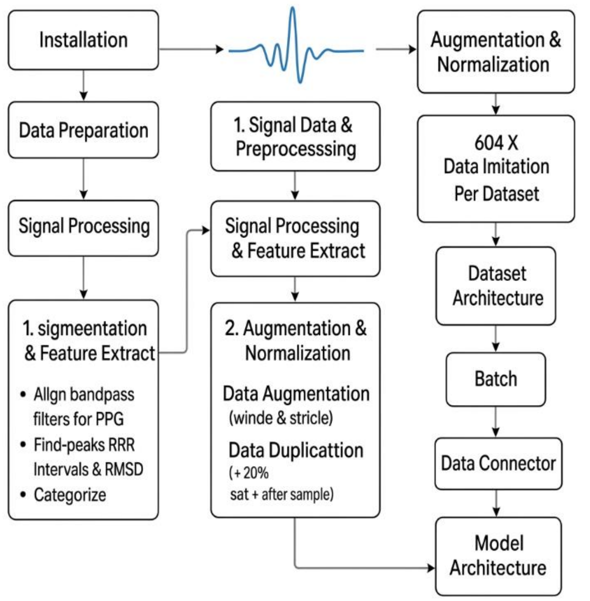

Figure 8. End-to-end deep learning pipeline for classifying biosignals (PPG and PCG)

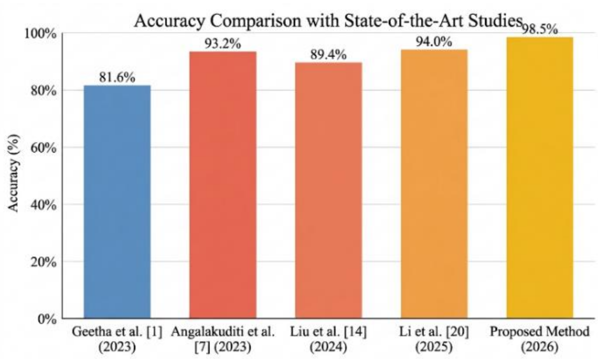

Figure 9. Performance comparison of the proposed PCG–PPG stress detection framework with recent state-of-the-art studies (2023–2025)

Authors' Contribution
All authors have contributed equally to prepare the paper.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Geetha R, et al. A Novel Deep Learning based Stress Analysis and Detection Scheme using Characteristic Data. In: 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). 2023.

[2] Durga CSLV, Manimaran J, Reddy MP. An Empirical Overview of Stress Detection Based on Photoplethysmography Signal Using Various Deep Learning Techniques. In: 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT). 2024.

[3] Tarun M, et al. Stress Detection by Deep Learning Technique. In: 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). 2024.

[4] Emadi M, Surani S. A New Method in Improving the Accuracy of Fetal Brain Health Diagnosis Based on Image Analysis of Constrictors Feature in Ultrasound Images. Int J Biophotonics Biomed Eng. 2025;5(1).
DOI: https://doi.org/10.71498/ijbbe.2025.1195960

[5] Benchekroun M, et al. Comparison of Stress Detection through ECG and PPG signals using a Random Forest-based Algorithm. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022.
DOI: https://doi.org/10.1109/EMBC48229.2022.9871906

[6] Eren E, Navruz TS. Stress Detection with Deep Learning Using BVP and EDA Signals. In: 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). 2022.

[7] Angalakuditi H, Bhowmik B. Stress Detection Using Deep Learning Algorithms. In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2023.

[8] Rajabioun M. Identifying Autism from EEG Signals Using Features Derived from Active Brain Source Models. Int J Biophotonics Biomed Eng. 2025;5(1):55-68.
DOI: https://doi.org/10.71498/ijbbe.2025.1204968

[9] Kyrou M, Kompatsiaris I, Petrantonakis PC. Deep Learning Approaches for Stress Detection: A Survey. IEEE Trans Affect Comput. 2025;16(2):499-517.
DOI: https://doi.org/10.1109/TAFFC.2022.3188749

[10] Abdelfattah E, Joshi S, Tiwari S. Machine and Deep Learning Models for Stress Detection Using Multimodal Physiological Data. IEEE Access. 2025;13:4597-4608.
DOI: https://doi.org/10.1109/ACCESS.2025.3141881

[11] Ghoshe DK, Munir NS, Debnath S. Mental Workload Assessment from ECG and PPG Signal Using Machine Learning Techniques. In: 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON). 2024.

[12] Pashazadeh M, et al. A review of the antibacterial properties of zinc oxide nanoparticles: synthesis, mechanism of action, and medical applications. Int J Biophotonics Biomed Eng. 2025;5(1).

DOI: https://doi.org/10.71498/ijbbe.2025.1202308

[13] Khanlari M, Dinparvar MJ. Design and Experimental Validation of a Portable Photometer for Non-Invasive Neonatal Jaundice Assessment. Int J Biophotonics Biomed Eng. 2025;5(1):69-78.
DOI: https://doi.org/10.71498/ijbbe.2025.1211116

[14] Liu Z, et al. Mental Stress Detection Using PPG Signals Based on Transformer-LSTM Model. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2024.

[15] Gonzalez-Vazquez JJ, et al. A Deep Learning Approach to Estimate Multi-Level Mental Stress From EEG Using Serious Games. IEEE J Biomed Health Inform. 2024;28(7):3965-72.
DOI: https://doi.org/10.1109/JBHI.2024.3378092

[16] Lazović A, Tadić P, Đorđević N, Atanasoski V, Tiosavljevic M, Ivanovic M, Hadzievski L, Ristic A, Vukcevic V, Petrovic J. SensSmartTech database of cardiovascular signals synchronously recorded by an electrocardiograph, phonocardiograph, photoplethysmograph and accelerometer (version 1.0.0). PhysioNet. 2024.
DOI: https://doi.org/10.13026/fy9p-n277

[17] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 2000;101(23):e215-e220.
DOI: https://doi.org/10.1161/01.CIR.101.23.e215

[18] Andalibi Miandoab S, Ghasemzadeh N. Computational modeling of M1-BG-Th network firing rate and beta oscillation in brain neurological diseases for treatment with electrical or optogenetic stimulation. Int J Biophotonics Biomed Eng. 2025;5(1).
DOI: https://doi.org/10.71498/ijbbe.2025.1191596

[19] Patel A, Nariani D, Rai A. Mental Stress Detection using EEG and Recurrent Deep Learning. In: 2023 IEEE Applied Sensing Conference (APSCON). 2023.
DOI:
https://doi.org/10.1109/APSCON56919.2023.10075003

[20] Aghamohammadian M, Vahedi A, Haghipour S. Optical biosensor for detection of hemoglobin using ternary photonic crystals. Int J Biophotonics Biomed Eng. 2025;5(1).
DOI: https://doi.org/10.71498/ijbbe.2025.1193386

[21] HG T S, et al. Prediction of Cardiovascular Disease from Retinal Images using Deep Learning. In: 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP). 2024.
DOI: https://doi.org/10.1109/AISP58276.2024.10564342

[22] Li M, et al. Stress Severity Detection in College Students Using Emotional Pulse Signals and Deep Learning. IEEE Trans Affect Comput. 2025:1-13.
DOI: https://doi.org/10.1109/TAFFC.2023.3287146

[23] Wang W, Najafizadeh L. Ultra-Short Term Heart Rate Variability Estimation Using PPG and End-to-End Deep Learning. In: 2024 58th Asilomar Conference on Signals, Systems, and Computers. 2024.
DOI:
https://doi.org/10.1109/IEEECONF57900.2024.1234567

[24] Kechris C, Delopoulos A. RMSSD Estimation From Photoplethysmography and Accelerometer Signals Using a Deep Convolutional Network. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2021.
DOI: https://doi.org/10.1109/EMBC46164.2021.9630724

[25] Khomidov M, Lee JH. The Novel Estimation Algorithm of Heart Rate Variability and Stress Using Facial Video Analysis. In: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2024.

[26] Mojtahed H, et al. A Deep Learning Approach for Heart Rate Variability Error Correction. In: 2024 IEEE 6th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). 2024.

[27] Asjad M, et al. Federated Focal Modulated UNet for Cardiovascular Image Segmentation. In: 2024 IEEE International Conference on Big Data (BigData). 2024.

[28] Guru RP, et al. Enhanced Cardiovascular Disease Detection from Phonocardiogram Signals Using Deep Learning & Wavelet-Based Denoising. In: 2024 4th Asian Conference on Innovation in Technology (ASIANCON). 2024.

[29] Meghana ML, et al. Cardiovascular Disease Detection in ECG Images Using CNN-Bi-LSTM Model. In: 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS). 2024.

[30] Sharma RK, et al. An Ensemble Convolutional Neural Network-Bidirectional Long Short-Term Memory-Attention Approach for Electrocardiogram-Based Heart Disease Diagnosis. In: 2025 International Conference on Microwave, Optical, and Communication Engineering (ICMOCE). 2025.

[31] Fayouka A, et al. Cardiac Segmentation: A Comparative Study Between 3D UNet and 2D UNet performances. In: 2024 IEEE/ACS 21st International Conference on Computer Systems and Applications (AICCSA). 2024.

[32] Novoselnik F, et al. 3D U-Net based method for fast segmentation of whole heart from CT images. In: 2022 International Symposium ELMAR. 2022.

[33] Sunilkumar G, Kumaresan P. Deep Learning and Transfer Learning in Cardiology: A Review of Cardiovascular Disease Prediction Models. IEEE Access. 2024;12:193365-86.
DOI: https://doi.org/10.1109/ACCESS.2024.3352501

[34] Hasanpoor Y, et al. Stress Detection Using PPG Signal and Combined Deep CNN-MLP Network. In: 2022 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME). 2022.

[35] Roy TS, et al. Valvular Heart Disease Prediction Using CNN-based Residual Network. In: 2023 16th International Conference on Sensing Technology (ICST). 2023.

[36] Kannan A, et al. Detection of Valvular Heart Diseases From PCG Signals Using Machine and Deep Learning Models: A Review. IEEE Access. 2025;13:110344-64. DOI: https://doi.org/10.1109/ACCESS.2025.3350124

[37] Reshan MSA, et al. A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks. IEEE Access. 2023;11:121574-91. DOI: https://doi.org/10.1109/ACCESS.2023.3325193

[38] Kumar A, et al. Optimized Almond Damage Detection with UNet and Xception-based Segmentation and Classification. In: 2025 International Conference on Electronics and Renewable Systems (ICEARS). 2025.

[39] Kotipalli KD, et al. Machine Learning and Deep Learning Analysis of PCG Data. In: 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP). 2024. DOI: https://doi.org/10.1109/AISP58276.2024.10564351

[40] Nandan S, Mandal S, Ghosal P. Stress Detection and Monitoring: A Systematic Review. In: 2024 IEEE International Symposium on Smart Electronic Systems (iSES). 2024.

[41] Sanchez OD, et al. Real-Time Neural Classifiers for Sensor Faults in Three Phase Induction Motors. IEEE Access. 2023;11:19657-68. DOI: https://doi.org/10.1109/ACCESS.2023.3279745

[42] Sengar A, et al. Hybrid Approach for Heart Disease Detection using Classification Algorithms. In: 2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG). 2023.

[43] Shruthi K, Naidu RCA. Deep Learning for Cardiovascular Disease Detection: Comprehensive study on CNN and Random Forest based Hybrid Approach. In: 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS). 2025.