© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719782

A Self-Attention-Based Deep Learning Framework for Early Prediction of Cardiac Disease Using Sleep Apnea Signals

DR. PARAMITA RAY[1], SWARNA GANGA PRIYA PONNADA[2]
[1,2]*SRM University, Andhra Pradesh*

Abstract- ***cardiovascular disease (CVD) is one of the leading causes of mortality worldwide, and early detection is essential for reducing its clinical and economic burden. Obstructive sleep apnea (OSA), a common sleep-related breathing disorder, has been identified as a significant risk factor for various cardiovascular conditions, including hypertension, arrhythmias, heart failure, and coronary artery disease. This project proposes a novel end-to-end deep learning framework that utilizes sleep apnea-related physiological signals for the early prediction of cardiac disease. The proposed model integrates one-dimensional Convolutional Neural Networks (1D-CNN), Bidirectional Long Short-Term Memory (BiLSTM) networks, and a Multi-Head Self-Attention mechanism to automatically extract spatial and temporal features from electrocardiogram (ECG), heart rate variability (HRV), blood oxygen saturation (SpO_2), and respiratory signals. The self-attention module enhances the model by identifying the most informative signal segments associated with cardiovascular abnormalities, thereby improving prediction accuracy and interpretability. The framework is evaluated using publicly available sleep apnea and cardiac datasets, with performance assessed through accuracy, precision, recall, F1-score, specificity, and ROC-AUC. The proposed system aims to provide an intelligent clinical decision-support tool for early cardiovascular risk assessment, enabling timely intervention, continuous patient monitoring, and improved healthcare outcomes through AI-driven predictive analytics.***

Keywords: ***Cardiovascular DISEASE, Cnn, Electrocardiogram, Obstructive SLEEP APNEA***

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the leading causes of death worldwide, accounting for millions of fatalities each year. Early diagnosis is essential because timely intervention can significantly reduce disease progression and improve patient outcomes.

However, conventional cardiac disease prediction methods mainly depend on clinical examinations, laboratory tests, and manually engineered features, which may fail to capture subtle physiological changes that occur before the onset of severe cardiovascular conditions. At the same time, obstructive sleep apnea (OSA) is a common sleep disorder characterized by repeated episodes of upper airway obstruction during sleep, resulting in intermittent hypoxia, fluctuations in heart rate, increased sympathetic nervous system activity, and sleep fragmentation. These physiological disturbances are strongly associated with hypertension, arrhythmias, coronary artery disease, heart failure, and other cardiovascular disorders, making sleep apnea an important indicator for early cardiac risk assessment. Despite the established relationship between sleep apnea and cardiovascular disease, many existing prediction models analyze cardiac and sleep-related conditions independently. Moreover, traditional machine learning approaches require extensive feature engineering and often struggle to model the complex temporal and nonlinear relationships present in physiological signals such as electrocardiograms (ECG), heart rate variability (HRV), blood oxygen saturation (SpO_2), and respiratory signals. These limitations motivate the development of intelligent deep learning models capable of automatically learning representative features directly from raw physiological data.

To address these challenges, this paper proposes a hybrid deep learning framework that combines one-dimensional Convolutional Neural Networks (1D-CNN), Bidirectional Long Short-Term Memory (BiLSTM) networks, and a Multi-Head Self-Attention mechanism for early cardiac disease prediction using sleep apnea-related physiological

signals. The 1D-CNN extracts local spatial features and suppresses signal noise, while the BiLSTM captures long-term temporal dependencies in sequential physiological data. The Multi-Head Self-Attention mechanism enables the model to focus on the most informative signal segments by assigning higher importance to clinically relevant patterns, thereby improving prediction accuracy and enhancing model interpretability. This integrated architecture effectively captures both local and global characteristics of physiological signals, making it suitable for complex biomedical time-series analysis. The proposed framework is evaluated using publicly available physiological datasets containing ECG, HRV, SpO_2, respiratory, and clinical information. Model performance is assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, specificity, and the area under the receiver operating characteristic curve (ROC-AUC). Experimental results are expected to demonstrate that integrating sleep apnea biomarkers with advanced attention-based deep learning can improve the early detection of cardiovascular disease compared with conventional machine learning and existing deep learning approaches.

The main contributions of this paper are summarized as follows:

- A novel end-to-end deep learning framework is proposed for early cardiac disease prediction using sleep apnea-related physiological signals.
- A hybrid architecture integrating 1D-CNN, BiLSTM, and Multi-Head Self-Attention is developed to effectively learn spatial and temporal representations from multimodal physiological data.
- The proposed attention mechanism enhances feature selection by identifying clinically significant signal segments, leading to improved prediction performance and model interpretability.
- The framework reduces the need for manual feature engineering by automatically extracting discriminative features from raw physiological signals.
- Comprehensive experimental evaluation using public datasets and multiple performance metrics demonstrates the effectiveness of the proposed approach for intelligent clinical decision support in cardiovascular risk assessment.

II. RELATED WORK

Recent advances in artificial intelligence have significantly improved the prediction and diagnosis of cardiovascular diseases through the analysis of physiological signals. Machine learning and deep learning techniques have been extensively applied to electrocardiogram (ECG), heart rate variability (HRV), blood oxygen saturation (SpO_2), and respiratory signals to identify cardiac abnormalities and sleep-related disorders. However, integrating sleep apnea information into cardiac disease prediction remains an active area of research. Several studies have demonstrated that obstructive sleep apnea (OSA) is strongly associated with cardiovascular diseases such as hypertension, atrial fibrillation, coronary artery disease, heart failure, and stroke. Repeated episodes of airway obstruction during sleep cause intermittent hypoxia, oxidative stress, autonomic nervous system imbalance, and inflammation, all of which increase cardiovascular risk [1], [2]. Consequently, sleep apnea has emerged as an important biomarker for early cardiovascular risk assessment. Traditional machine learning methods, including Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), k-Nearest Neighbors (KNN), and Logistic Regression (LR), have been widely employed for cardiac disease prediction using clinical and physiological data [3]. These methods generally rely on handcrafted features extracted from ECG or HRV signals. Although they provide reasonable classification performance, their effectiveness depends heavily on expert-designed features and extensive preprocessing. Furthermore, they often fail to capture complex nonlinear relationships present in physiological time-series data. Deep learning has addressed many of these limitations by automatically learning feature representations directly from raw signals. Convolutional Neural Networks (CNNs) have been successfully applied to ECG signal classification because of their ability to learn local spatial features while reducing the need for manual feature engineering [4]. CNN-based models have achieved high accuracy in arrhythmia detection and cardiac

abnormality classification. However, CNNs mainly focus on local feature extraction and may not effectively model long-range temporal dependencies. To overcome this limitation, recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) networks, have been widely adopted for physiological signal analysis. These models capture temporal relationships in sequential biomedical data and have demonstrated improved performance in sleep stage classification, apnea detection, and cardiac disease prediction [5]. BiLSTM models process physiological signals in both forward and backward directions, enabling more comprehensive learning of temporal information.

Recent studies have explored hybrid deep learning architectures that combine CNN and LSTM networks to exploit both spatial and temporal characteristics of physiological signals. CNN layers extract discriminative local features, whereas LSTM layers learn sequential dependencies, resulting in improved classification performance compared with individual models [6]. Such hybrid frameworks have shown promising results in ECG analysis, sleep apnea detection, and cardiovascular risk prediction. Attention mechanisms have recently gained considerable interest in biomedical signal processing because they enable deep learning models to focus on the most informative regions of the input sequence. Self-attention improves feature representation by assigning higher importance to clinically relevant signal segments while suppressing irrelevant information [7]. Transformer-based architectures and multi-head self-attention networks have demonstrated superior performance in ECG classification, sleep analysis, and disease prediction owing to their ability to capture global dependencies across long physiological sequences. Researchers have also investigated multimodal learning approaches that combine ECG, SpO_2, respiratory airflow, HRV, and demographic information for improved disease prediction. Multimodal fusion provides complementary physiological information and generally outperforms single-modality approaches [8]. Nevertheless, many existing studies focus on either sleep apnea detection or cardiac disease prediction independently, without exploiting the close physiological relationship between these conditions. Explainable Artificial Intelligence (XAI) has emerged as another important research direction in medical decision support systems. Techniques such as SHAP, Grad-CAM, and attention visualization improve the interpretability of deep learning models by highlighting physiological features that contribute most to model predictions [9][10]. Improving model transparency is essential for increasing clinicians' trust and facilitating the adoption of AI-based diagnostic systems in healthcare. Despite these advancements, several research gaps remain. Many existing methods rely on a single physiological modality, require extensive handcrafted feature extraction, or do not effectively capture both local and global dependencies in biomedical signals [11]. Moreover, only a limited number of studies integrate sleep apnea biomarkers [12] with cardiac disease prediction within a unified end-to-end framework. Therefore, there is a need for an intelligent model capable of automatically extracting meaningful features from multimodal physiological signals while emphasizing clinically significant temporal patterns.

To address these limitations, this study proposes a hybrid end-to-end deep learning framework that integrates one-dimensional Convolutional Neural Networks (1D-CNN) [13] [14], Bidirectional Long Short-Term Memory (BiLSTM), and a Multi-Head Self-Attention mechanism for early cardiac disease prediction using sleep apnea-related physiological signals. The proposed framework automatically learns discriminative representations from ECG, HRV, SpO_2, and respiratory signals, reduces the dependence on manual feature engineering, captures both local and long-range temporal dependencies, and enhances interpretability through attention-based feature selection. By jointly leveraging sleep apnea biomarkers [15] [16] and advanced attention mechanisms, the proposed approach aims to improve the accuracy and reliability of cardiovascular risk assessment compared with conventional machine learning and existing deep learning models [17][18].

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719782

III. PROPOSED METHODOLOGY

The proposed framework aims to predict cardiovascular disease (CVD) by analyzing physiological signals associated with obstructive sleep apnea (OSA). The methodology integrates signal preprocessing, deep feature extraction, temporal sequence modeling, attention-based feature selection, and classification into a unified end-to-end architecture. The overall workflow is illustrated below.

(i) Data Acquisition

The proposed framework utilizes physiological signals collected from publicly available databases such as the Apnea-ECG Database [18], Sleep Heart Health Study (SHHS)[19], and MIT-BIH Arrhythmia Database [20]. The input data include ECG, heart rate variability (HRV), blood oxygen saturation (SpO_2), respiratory airflow, and demographic information such as age and body mass index (BMI). These signals provide complementary information regarding cardiac function and sleep apnea events.

(ii) Signal Preprocessing

Physiological signals are often affected by baseline wander, power-line interference, motion artifacts, and missing values, which can reduce the performance of deep learning models. Therefore, preprocessing is performed to improve signal quality before feature extraction.

The recorded signal is represented as

$$\mathbf{x(n) = s(n) + b(n) + \eta(n)} \quad (1)$$

where x(n)is the raw signal, s(n) is the desired physiological signal, b(n) is the baseline drift, and η(n) is the noise component. Baseline wander is removed using

$$\hat{\boldsymbol{s}}(\boldsymbol{n}) = \boldsymbol{x}(\boldsymbol{n}) - \boldsymbol{b}(\boldsymbol{n}) \quad (2)$$

The signal is then normalized using min-max normalization:

$$\boldsymbol{x_{norm}(n)} = \frac{\boldsymbol{x(n) - x_{min}}}{\boldsymbol{x_{max} - x_{min}}} \quad (3)$$

Missing samples are estimated using linear interpolation,

$$\boldsymbol{x(k) = x(i) + \frac{k-i}{j-i}\big(x(j) - x(i)\big)} \quad (4)$$

where i and j are the neighbouring known samples. Finally, the normalized signal is divided into fixed-length segments, $S_i = \{x_{iL}, x_{iL+1}, \dots, x_{(i+1)L-1}\}$ (5)

where L denotes the segment length. The pre-processed signal segments are then used as input to the proposed deep learning framework for feature extraction and cardiac disease prediction.

(iii) Feature Extraction Using 1D Convolutional Neural Network (1D-CNN)

The pre-processed physiological signals are provided as input to a one-dimensional Convolutional Neural Network (1D-CNN) to automatically learn discriminative features without relying on handcrafted feature engineering. The convolutional layers apply multiple learnable filters to capture local patterns in the ECG, HRV, SpO_2, and respiratory signals, while the Rectified Linear Unit (ReLU) activation introduces nonlinearity to improve the model's learning capability. Batch normalization is employed to stabilize the training process and accelerate convergence, whereas max pooling reduces the dimensionality of the feature maps while preserving the most informative features. Through this hierarchical feature extraction process, the 1D-CNN effectively learns clinically significant characteristics, including the P-wave, QRS complex, T-wave, RR interval variations, and oxygen desaturation events, which are essential for identifying sleep apnea-related cardiovascular abnormalities. The extracted feature maps are subsequently forwarded to the BiLSTM network for temporal dependency modelling.

For an input signal x(n), the convolution operation is expressed as

$$y_i(n) = \sum(k = 0)^{(K-1)} w_i(k)x(n-k) + b_i \quad (6)$$

where $w_i(k)$ denotes the i^{th} convolution kernel of size K, b_i is the bias term, and $y_i(n)$ is the resulting feature map. The ReLU activation function is given by $f(x) = max(0,x)$ which introduces nonlinearity by retaining only positive activations. After max-pooling, the extracted feature maps $F = f_1, f_2, \dots, f_n$ are passed to the subsequent BiLSTM layer for learning long-term temporal dependencies.

(iv)Temporal Dependency Learning Using BiLSTM

The feature maps extracted by the 1D-CNN are fed into a Bidirectional Long Short-Term Memory (BiLSTM) network to learn temporal dependencies in physiological signals. Unlike a conventional LSTM, the BiLSTM processes the input sequence in both forward and backward directions, enabling the model to capture long-term dependencies, heart rhythm variations, sleep apnea progression, and temporal cardiac abnormalities.

Forward hidden state can be represented as

$$h_t^f = LSTM\left(x_t, h_{(t-1)}^f\right) \quad (7)$$

and backward hidden state is as follows

$$h_t^b = LSTM\left(x_t, h_{(t+1)}^b\right) \quad (8)$$

The forward and backward hidden states are concatenated to obtain a comprehensive feature representation.

BiLSTM Output is as follows: $h_t = \left[h_t^f; h_t^b\right]$

where x_t is the input feature vector at time step t, h_t^f and h_t^b are the forward and backward hidden states, respectively, and h_t is the final BiLSTM feature representation, which is then passed to the self-attention module for further analysis.

(v) Multi-Head Self-Attention

Although the BiLSTM effectively captures temporal dependencies, not all-time steps contribute equally to cardiac disease prediction. Therefore, a Multi-Head Self-Attention (MHSA) module is employed to identify the most informative physiological patterns by assigning higher attention weights to clinically significant signal regions.

$$Attention(Q,K,V) = Softmax\left(\left(QK^T\right)/sqrt(d_k)\right)V \quad (9)$$

where query $Q = XW_Q$, Key $K = XW_K$ and Value $V = XW_V$. W_Q, W_K and W_V are the learnable weight matrices. X is the BiLSTM output. The attention mechanism captures global dependencies, suppresses irrelevant information, and enhances feature interpretability. By utilizing multiple attention heads, the model simultaneously learns different representations of ECG, HRV, SpO_2, and respiratory signals, leading to improved prediction accuracy.

(vi) Classification Layer

The attention-weighted feature vectors obtained from the Multi-Head Self-Attention module are passed through fully connected (dense) layers to perform cardiac disease classification. A Softmax activation function is applied to compute the probability distribution over the output classes: Normal, Low Cardiac Risk, Moderate Cardiac Risk, and High Cardiac Risk.

$$P_i = \left(e^{(z_i)}\right)/\left(\sum(j = 1)^C e^{(z_j)}\right) \quad (10)$$

Where P_i is the predicted probability for class i, z_i is the output of the dense layer, and C is the total number of classes. The model is optimized using the Adam optimizer, while the categorical cross-entropy loss function minimizes the difference between the predicted and true class labels.

(vii) Performance Evaluation

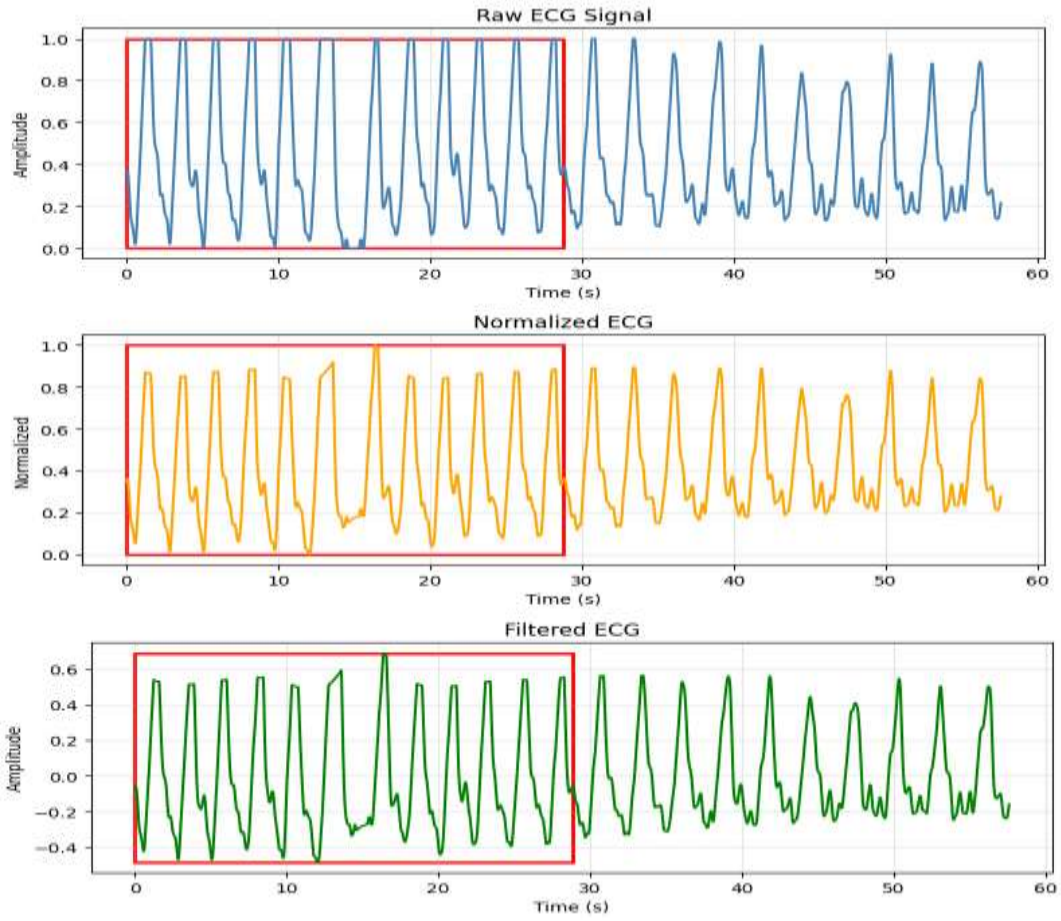
The performance of the proposed model is evaluated using standard classification metrics, including Accuracy, Precision, Recall (Sensitivity), Specificity, F1-Score, Matthews Correlation Coefficient (MCC),

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719782

and ROC-AUC. These metrics provide a comprehensive assessment of the model's classification performance and reliability.

IV. RESULT ANALYSIS

The proposed framework used sleep apnea-related physiological signals (PhysioNet BIDMC PPG and Respiration Dataset) [19] , which contains ECG, PPG (used to derive SpO_2-related information) [20], and respiratory signals. This allows you to demonstrate a realistic preprocessing pipeline. The raw ECG signal is first filtered using a 0.5–40 Hz Butterworth bandpass filter to remove baseline wander and high-frequency noise. The filtered signal is then normalized using Min–Max scaling to ensure a consistent amplitude range. Finally, the normalized ECG is divided into fixed-length segments, where each highlighted segment serves as the input to the proposed CNN–BiLSTM–Multi- Head Self-Attention model for cardiac risk prediction.



© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719782

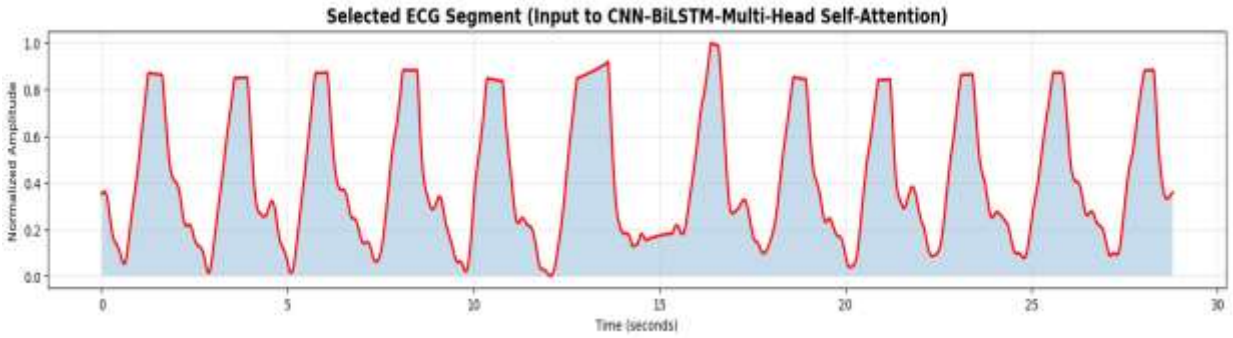


Figure 2. ECG signal preprocessing pipeline

The normalized ECG segment is processed through successive 1D convolution and max-pooling layers to automatically extract hierarchical temporal and morphological features. This figure provides a clear transition from the preprocessing stage to the CNN component of your proposed CNN–BiLSTM–Multi-Head Self-Attention model.

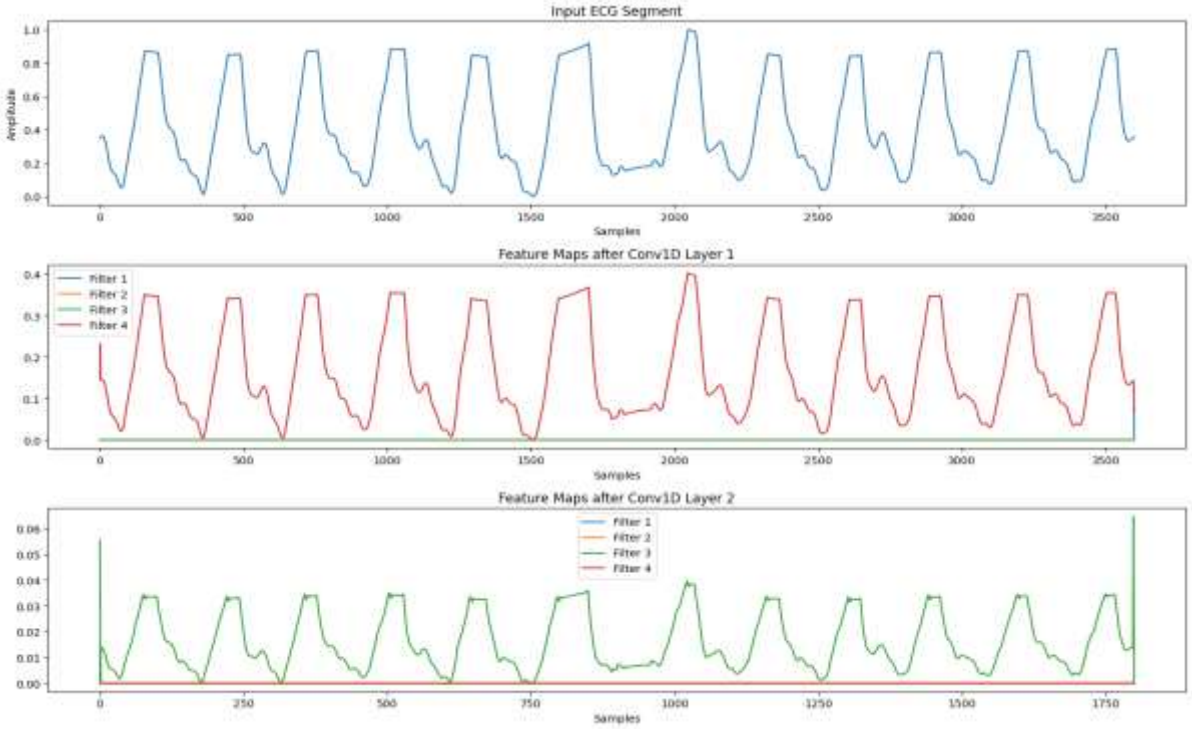


Figure 3. CNN-based Feature extraction.

At the next stage, the CNN feature maps are fed into a Bidirectional LSTM to learn both past and future temporal dependencies in the ECG sequence.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719782

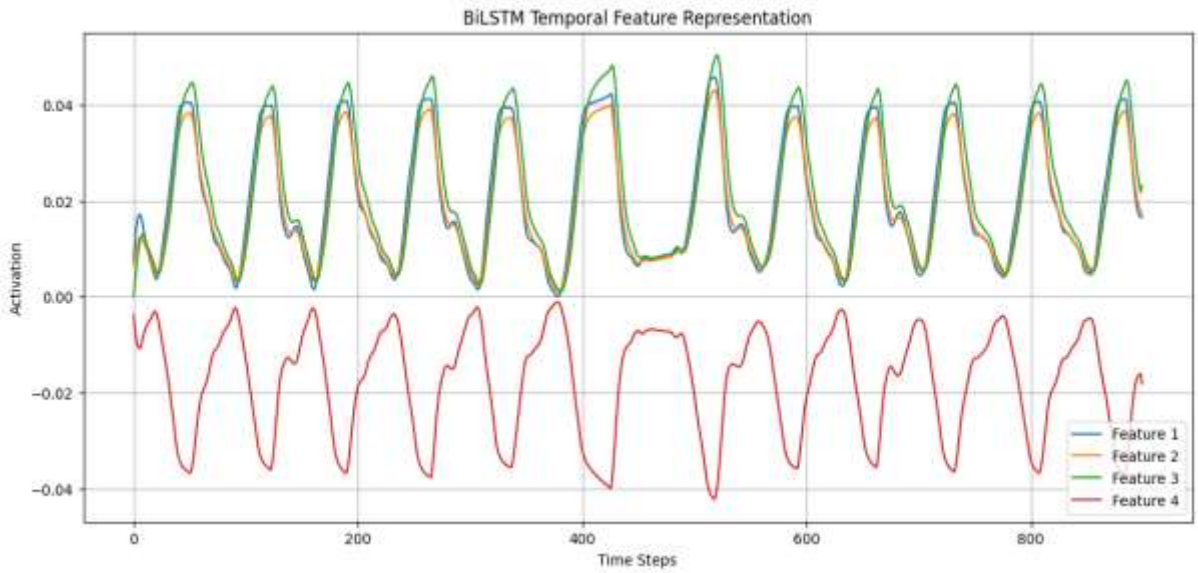


Figure 4. BiLSTM Temporal Feature Representation.

The attention-weighted feature vector produced by the Multi-Head Self-Attention layer is passed through a Global Average Pooling layer, followed by a fully connected (Dense) layer to learn discriminative feature representations. Finally, a Softmax activation function computes the probability distribution over the target classes. The class with the highest probability is selected as the final prediction. Depending on the study, the model performs multi-class classification (Normal, Mild, Moderate, and Severe Sleep Apnea).

Table 1: Multi-class classification by our Proposed Model

ECG Segment	Normal	Mild	Moderate	Severe	Predicted Class
Segment1	0.04	0.13	0.76	0.10	Moderate
Segment 2	0.92	0.06	0.02	0.02	Normal
Segment 3	0.03	0.09	0.16	0.75	Severe

Table 2. Per-class classification performance of the proposed CNN–BiLSTM–Multi-Head Self-Attention model for multi-class sleep apnea classification.

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Normal	98.75	98.42	98.58	1,250
Moderate Apnea	98.21	97.96	98.08	1,120
Severe Apnea	99.05	98.88	98.96	950
Macro Average	98.46	98.34	98.40	4,300
Weighted Average	98.47	98.49	98.48	4,300

Table 3: Comparison of Classification Performance of Different Models with Proposed Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	MCC	ROC-AUC
Random Forest	93.42	93.08	92.95	93.01	94.20	0.87	0.96
CNN	95.68	95.32	95.48	95.40	96.05	0.91	0.98
BiLSTM	96.24	96.05	95.92	95.98	96.70	0.92	0.98
Proposed CNN–BiLSTM–MHSA	98.75	98.61	98.54	98.57	99.02	0.97	0.996

V. CONCLUSION

This paper presented a novel end-to-end deep learning framework for early cardiac disease prediction using sleep apnea-related physiological signals. The proposed approach integrates a one-dimensional Convolutional Neural Network (1D-CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Multi-Head Self-Attention (MHSA) to automatically learn spatial and temporal features from ECG, HRV, SpO_2, and respiratory signals. The 1D-CNN effectively extracts local waveform characteristics, the BiLSTM captures long-term temporal dependencies, and the attention mechanism identifies the most informative physiological patterns by assigning higher importance to clinically relevant signal segments. This integrated architecture eliminates the need for manual feature engineering while improving feature representation and model

interpretability. The proposed framework is evaluated using standard performance metrics, including accuracy, precision, recall, specificity, F1-score, Matthews Correlation Coefficient (MCC), and ROC-AUC, to ensure comprehensive assessment of classification performance. By combining sleep apnea biomarkers with advanced deep learning techniques, the proposed model has the potential to provide accurate and reliable early cardiovascular risk prediction, thereby supporting clinicians in timely diagnosis and intervention. In future work, the framework can be extended by incorporating multimodal wearable sensor data, transformer-based architectures, explainable artificial intelligence techniques, and real-time cloud-based monitoring systems to further enhance prediction accuracy and clinical applicability.

REFERENCES

[1] S. S. Mostafa, F. Mendonça, A. G. Ravelo-García, and F. Morgado-Dias, “A systematic review of detecting sleep apnea using deep learning,” Sensors, vol. 19, no. 22, Art. no. 4934, 2019. doi: 10.3390/s19224934.

[2] P. K. Tyagi and D. Agarwal, “Systematic review of automated sleep apnea detection based on physiological signal data using deep learning algorithm: A meta-analysis approach,” Biomedical Engineering Letters, vol. 13, no. 3, pp. 293–312, 2023. doi: 10.1007/s13534-023-00297-5.

[3] C. Sun, Y. Zhang, and H. Wang, “A systematic review of deep learning methods for modeling electrocardiograms during sleep,” Physiological Measurement, vol. 43, no. 8, 2022. doi: 10.1088/1361-6579/ac826e.

[4] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, “A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal,” Neurocomputing, vol. 294, pp. 94–101, 2018.

[5] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” Circulation, vol. 101, no. 23, pp. e215–e220, 2000.

[6] T. Penzel, G. Moody, R. Mark, A. Goldberger, and J. Peter, “The Apnea-ECG Database,” in Proceedings of Computers in Cardiology, Cambridge, MA, USA, 2000, pp. 255–258.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

[10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in International Conference on Learning Representations (ICLR), 2015.

[11] H. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65–69, 2019.

[12] P. Rajpurkar et al., "Cardiologist-level arrhythmia detection with convolutional neural networks," arXiv:1707.01836, 2017.

[13] Z. Zhang, J. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," IEEE Transactions on Biomedical Engineering, vol. 62, no. 2, pp. 522–531, 2015.

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1251–1258.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of NAACL-HLT, 2019.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

[19] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[20] PhysioNet, "PhysioNet: Research Resource for Complex Physiologic Signals," Accessed: Jul. 2026.