Global Journal of Engineering and Technology Advances

eISSN: 2582-5003

Cross Ref DOI: 10.30574/gjeta

Journal homepage: https://gjeta.com/

(Research Article)

A Hybrid CNN–BiLSTM–LSTM–Attention Framework for Multichannel ECG-Based Cardiac Arrhythmia Classification

Nguyen Thi Bich Ngoc *

Faculty of Information Technology, SaoDo University, Hai Phong, Viet Nam.

Global Journal of Engineering and Technology Advances, 2026, 27(03), 045-052

Publication history: Received on 26 April 2026; revised on 02 June 2026; accepted on 04 June 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.27.3.0147

Abstract

This paper proposes a deep learning–based approach for cardiac arrhythmia classification from multichannel electrocardiogram (ECG) signals using a hybrid architecture that integrates Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), Long Short-Term Memory (LSTM), and a Multi-Head Self-Attention mechanism. The input data consist of four-channel ECG signals stored in CSV format and continuously recorded over time. Prior to model training, the ECG signals undergo a comprehensive preprocessing pipeline, including amplitude clipping within ±150 μV to remove impulsive artifacts, a 50 Hz notch filter for power-line interference suppression, a 0.5–45 Hz bandpass filter to preserve the physiological frequency components of ECG signals, and segment-wise normalization. After preprocessing, the signals are segmented into 4-second windows with 50% overlap, generating input samples of size (1024, 4). The segmented ECG windows are first processed by a one-dimensional CNN to extract local morphological features. Subsequently, BiLSTM and LSTM layers are employed to learn temporal dependencies within the signal sequences. The Multi-Head Self-Attention mechanism enables the model to focus on the most informative temporal regions of each ECG segment. Finally, a classification module consisting of Global Average Pooling, Dense, and Softmax layers performs heartbeat classification. Experimental results demonstrate that the proposed model achieves high classification performance in terms of Precision, Recall, and F1-score on the test dataset. These findings indicate that the combination of an effective ECG preprocessing strategy and a CNN–BiLSTM–LSTM–Attention architecture provides a promising solution for multichannel ECG-based arrhythmia classification..

Keywords: Electrocardiogram (ECG); Arrhythmia Classification; Deep Learning; CNN; BiLSTM; LSTM; Multi-Head Self-Attention

1. Introduction

Cardiovascular diseases remain one of the leading causes of mortality worldwide. Therefore, the early detection of abnormalities in cardiac activity plays a crucial role in timely diagnosis and treatment, thereby reducing the risk of complications and mortality. In clinical practice, electrocardiography (ECG) is a widely used, cost-effective, and noninvasive technique that provides valuable information regarding the electrical activity of the heart. However, ECG interpretation largely depends on physician expertise and can be adversely affected by signal noise, baseline drift, and motion artifacts caused by electrode contact disturbances.

In recent years, deep learning has been extensively applied in healthcare, particularly in the analysis of biomedical signals [1]. For ECG analysis, deep learning techniques have enabled more accurate and automated detection and classification of cardiac arrhythmias. Pan and Tompkins [2] introduced a QRS detection algorithm together with preprocessing techniques such as notch and bandpass filtering to improve signal quality. Johnson and Khoshgoftaar [3] addressed the issue of class imbalance and proposed class-weighting strategies to mitigate bias toward majority classes.

* Corresponding author: Nguyen Thi Bich Ngoc

Yin et al. [4] presented a CNN–BiLSTM–Attention architecture that combines local feature extraction, temporal dependency modeling, and attention mechanisms, serving as the foundation of the model proposed in this study. Furthermore, Yin et al. [5] provided a comprehensive review of deep learning approaches for ECG classification, highlighting the effectiveness of CNNs in extracting morphological ECG features. Finally, Vaswani et al. [6] introduced the Self-Attention and Multi-Head Attention mechanisms, which enable models to focus on the most informative regions of sequential data.

In this study, multichannel ECG signals stored in CSV format and continuously recorded over time are utilized. Since the raw ECG recordings are typically lengthy and contaminated by various sources of noise, a preprocessing framework is proposed that includes amplitude clipping at ±150 μV, 50 Hz notch filtering, 0.5–45 Hz bandpass filtering, and signal normalization. The preprocessed signals are then segmented into 4-second windows with 50% overlap to generate training samples for the deep learning model.

Based on this preprocessing framework, a hybrid deep learning architecture integrating CNN, BiLSTM, LSTM, and Multi-Head Self-Attention is proposed. The CNN component is responsible for extracting local morphological characteristics of ECG waveforms, whereas the BiLSTM and LSTM layers capture temporal dependencies within the signal sequences. The attention mechanism further enables the model to concentrate on the most informative signal regions for classification. The overall processing framework of the proposed approach is illustrated in Figure 1.

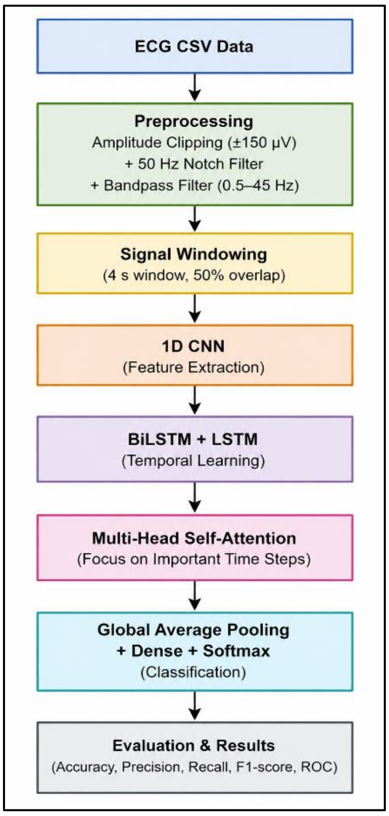

Figure 1 Overview of the proposed processing framework

2. Methodology

2.1. Overview of the Proposed Method

In this study, an ECG classification framework is developed based on a hybrid deep learning architecture that integrates CNN, BiLSTM, LSTM, and a Multi-Head Self-Attention mechanism. Unlike approaches that transform ECG signals into time–frequency representations, the proposed model learns discriminative features directly from preprocessed ECG signals in the time domain.

The CNN module is employed to extract local morphological characteristics from ECG waveforms across multiple channels. Subsequently, BiLSTM and LSTM layers are utilized to capture temporal dependencies within each signal segment. The Multi-Head Self-Attention mechanism enables the model to focus on the most informative temporal regions, thereby enhancing discrimination among different cardiac rhythm categories. Finally, a classification block consisting of Global Average Pooling, Dense, and Softmax layers produces the predicted class labels..

2.2. Input Data

The input data consist of multichannel ECG signals (four channels) stored in CSV format and continuously recorded over time. Due to the considerable length of the recordings, the raw signals are not directly used for training but are first segmented into fixed-length temporal windows.

Following preprocessing, the ECG signals are divided into 4-second segments with a sampling frequency of 128 Hz, corresponding to 1024 samples per segment. Adjacent windows overlap by 50%, which increases the number of training samples while preserving signal continuity. Consequently, each input sample fed into the model has a dimension of (1024, 4)..

2.3. Signal Preprocessing

Prior to segmentation, ECG signals undergo a preprocessing pipeline to suppress noise and improve signal stability. The preprocessing procedure is based on the framework proposed by Pan and Tompkins [2] and includes the following steps:

- Amplitude clipping within ±150 μV to eliminate impulsive artifacts.
- Application of a 50 Hz notch filter to remove power-line interference.
- Application of a 0.5–45 Hz bandpass filter to preserve the physiological ECG frequency range.
- Normalization of each signal segment to zero mean and unit variance.

This preprocessing strategy significantly improves signal quality and provides more reliable inputs for deep learning-based feature extraction and classification.

2.4. CNN-Based Feature Extraction Module

The preprocessed ECG segments are fed into a series of one-dimensional convolutional layers (Conv1D). CNNs have been demonstrated to be highly effective in automatically learning important morphological features of ECG waveforms, such as QRS complexes, P waves, and T waves, without requiring handcrafted feature engineering [7].

Each CNN block consists of:

- A one-dimensional convolutional layer (Conv1D),
- A Batch Normalization layer [8],
- A MaxPooling1D layer.

Batch Normalization stabilizes the training process by reducing internal covariate shift and allows the use of higher learning rates. The number of convolutional filters gradually increases throughout the network (32 → 64 → 128), enabling the extraction of increasingly complex hierarchical features. Meanwhile, MaxPooling progressively reduces the temporal dimension (1024 → 512 → 256 → 128), lowering computational complexity and mitigating overfitting.

2.5. Temporal Dependency Learning Using BiLSTM, LSTM, and Attention

The feature maps generated by the CNN module are subsequently processed by a Bidirectional Long Short-Term Memory (BiLSTM) layer.

The LSTM architecture, originally proposed by Hochreiter and Schmidhuber [9], effectively addresses the vanishing-gradient problem encountered in conventional recurrent neural networks through its gating mechanisms, including the forget gate, input gate, and output gate. BiLSTM extends the conventional LSTM by processing sequential information in both forward (past-to-present) and backward (future-to-present) directions, thereby capturing more comprehensive temporal dependencies.

Following the BiLSTM layer, an additional LSTM layer is employed to further refine temporal representations and enrich contextual information. The resulting feature sequence is then processed by a Multi-Head Self-Attention mechanism, which enables the model to identify and emphasize the most informative temporal regions within each ECG segment. This attention-based strategy improves the model's ability to distinguish among different cardiac rhythm patterns [6].

2.6. Classification Layer

After the attention module, the extracted features are passed through a Global Average Pooling (GAP) layer to reduce the number of trainable parameters and improve generalization performance. The final classification stage consists of a fully connected Dense layer followed by a Softmax activation function, which estimates the probability distribution across the target heartbeat classes and assigns the class with the highest probability as the prediction. The model is trained using the Sparse Categorical Cross-Entropy loss function and the Adam optimization algorithm [10] with a learning rate of 0.0001. Adam combines the advantages of Momentum and RMSProp by adaptively adjusting the learning rate for each parameter, making it particularly suitable for noisy biomedical signal data.

The training configuration is summarized as follows:

- Batch size: 32
- Number of epochs: 100
- Validation split: 0.2
- Early stopping patience: 8 epochs
- Class weighting strategy for handling class imbalance

Model performance is evaluated using Precision, Recall, and F1-score on the test dataset, providing a comprehensive assessment of classification effectiveness for multichannel ECG-based arrhythmia detection.

3. Experimental Results

3.1. Dataset Description

The experiments were conducted using the ECG Heartbeat Categorization Dataset [11], which contains electrocardiogram (ECG) recordings stored in CSV format. The dataset includes a variety of cardiac rhythm conditions, covering both normal heartbeats and common arrhythmia types.

Prior to model training, the ECG signals underwent a preprocessing procedure to suppress noise and standardize signal characteristics. Specifically, the DC component was removed by subtracting the mean value of each signal. Subsequently, a bandpass filter was applied to preserve the most relevant ECG frequency components, while a 50 Hz notch filter was used to eliminate power-line interference. After noise removal, the signals were segmented into fixed-length windows to ensure consistency of the model input.

In this study, the ECG recordings were categorized into four rhythm classes:

- Atrial Fibrillation (AFIB)
- Sinus Bradycardia (SB)
- Sinus Rhythm (SR)
- Sinus Tachycardia (ST)

The entire dataset was divided into training and testing subsets using an 80:20 ratio. This partitioning strategy ensures that the model has sufficient data for learning representative ECG features while enabling objective evaluation of its generalization capability on unseen samples.

3.2. Experimental Results

The training performance of the proposed model is illustrated through the Accuracy and Loss curves shown in Figures 2 and 3. The accuracy curves indicate rapid convergence of the proposed architecture. Training accuracy increases substantially during the initial epochs and gradually approaches nearly 100%. Similarly, test accuracy reaches a high level early in the training process and remains stable within the range of approximately 97–99%. The best validation performance is achieved around epoch 98, with a validation accuracy of approximately 97.1%, demonstrating strong generalization capability. The loss curves further confirm the effectiveness of the training process. Training loss decreases rapidly and approaches zero as the number of epochs increases. Meanwhile, validation loss drops significantly during the early training stage and subsequently stabilizes around 0.12 with only minor fluctuations. The absence of a sharp increase in validation loss indicates that severe overfitting does not occur and that the model maintains stable classification performance.

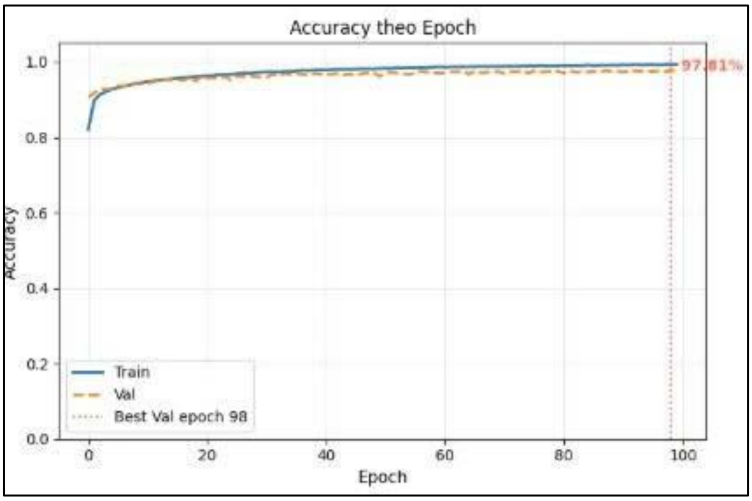

Figure 2 Accuracy curves of the proposed model

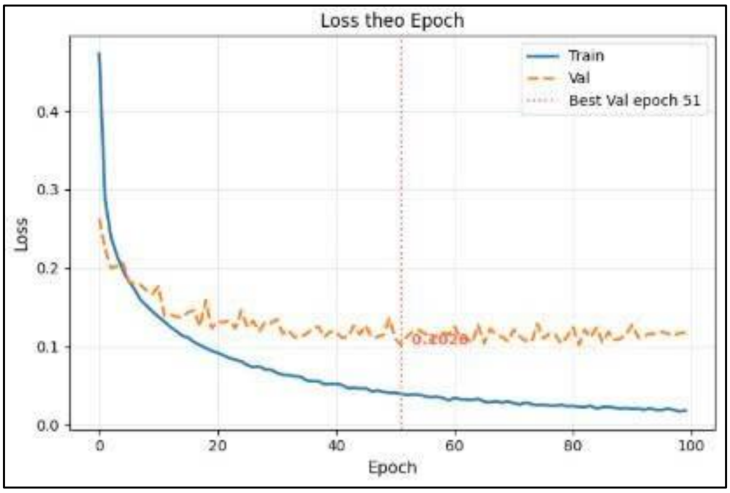

Figure 3 Loss curves of the proposed model

Based on the confusion matrix obtained from the test dataset consisting of 22,864 samples, the proposed model achieved an overall classification accuracy of 97.74%. The confusion matrix is a 4 × 4 matrix corresponding to the four cardiac rhythm classes: AFIB, SB, SR, and ST. Most correctly classified samples are concentrated along the main diagonal, indicating excellent classification performance.

Specifically, the AFIB class achieved a classification accuracy of 98.4%, with 4,099 correctly identified samples. The SB class attained an accuracy of 97.1%, with 8,402 correctly classified samples and only a small number misclassified as SR. The SR class achieved a classification accuracy of 97.4%, corresponding to 6,386 correctly predicted samples. The ST class exhibited the highest classification performance, with 99.1% of samples correctly recognized, corresponding to 3,460 instances. Most classification errors occurred between rhythm classes exhibiting similar ECG morphological characteristics, particularly SB and SR. However, the number of such misclassifications remained relatively small and had minimal impact on the overall reliability of the model. The confusion matrix is presented in Figure 4.

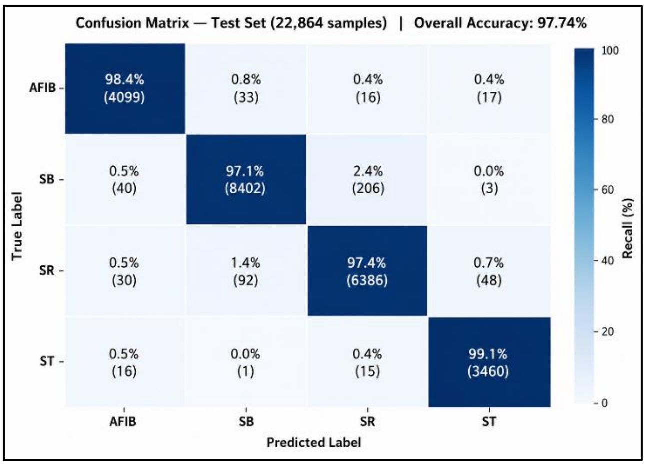

Figure 4 Confusion matrix of the proposed model

In addition to the confusion matrix, the multiclass classification capability of the proposed architecture was comprehensively evaluated using Receiver Operating Characteristic (ROC) analysis based on the One-vs-Rest strategy, as illustrated in Figure 5. The ROC curve depicts the relationship between the true positive rate (sensitivity) and the false positive rate (1 − specificity) under varying decision thresholds.

As shown in Figure 5, the ROC curves for all four cardiac rhythm classes closely approach the upper-left corner of the plot, indicating excellent discriminative capability. The macro-average Area Under the Curve (AUC) reached 0.999, demonstrating near-perfect classification performance. Notably, the ST class achieved an AUC value of 1.000, indicating perfect separation between sinus tachycardia and the remaining rhythm classes. Similarly, the AFIB, SB, and SR classes achieved remarkably high AUC values of 0.999, 0.998, and 0.997, respectively.

These results further demonstrate the robustness and reliability of the proposed CNN–BiLSTM–Attention architecture in recognizing complex ECG patterns, even when different rhythm classes exhibit substantial morphological similarities.

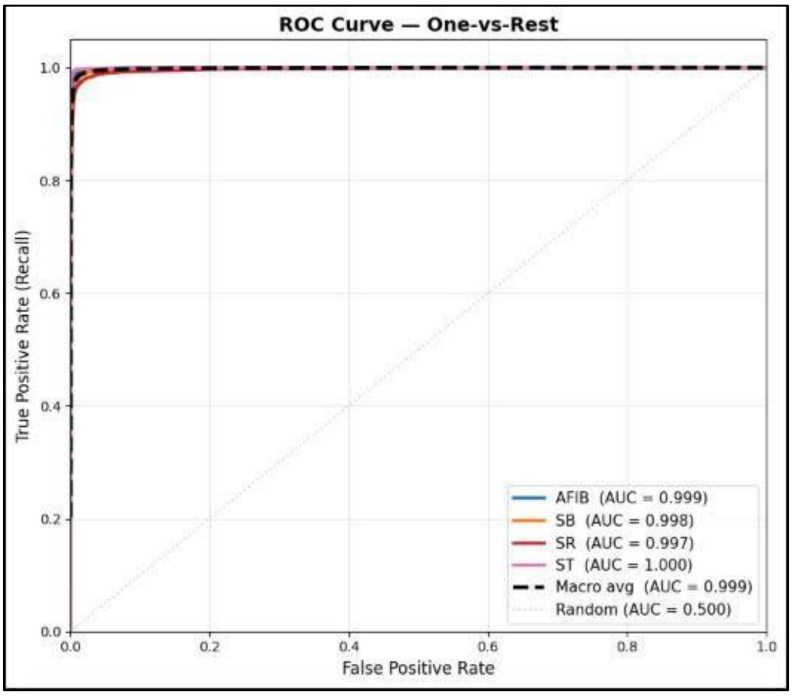

Figure 5 ROC curves of the proposed model

To further evaluate the effectiveness of the proposed method, four widely used performance metrics—Accuracy, Precision, Recall, and F1-score—were employed. Precision measures the proportion of correctly predicted samples among all samples assigned to a given class. Recall quantifies the model's ability to correctly identify samples that truly belong to that class. F1-score represents the harmonic mean of Precision and Recall and reflects the balance between these two metrics.

The experimental results indicate that the proposed CNN–BiLSTM–Attention model outperforms competing architectures, including ResNet-1D and Temporal Convolutional Networks (TCN), demonstrating superior classification accuracy and stability on ECG data. Furthermore, the proposed model maintains an excellent balance between Precision and Recall, resulting in a higher F1-score and improved overall classification performance. Detailed results are summarized in Table 1.

Table 1 Performance comparison of different models

Model	Accuracy	Precision	Recall	F1-Score
CNN–BiLSTM–Attention	97.74%	97.74%	98.01%	97.87%
ResNet-1D	93.40%	92.80%	93.10%	92.95%
TCN	91.20%	90.60%	99.00%	90.80%

Although the TCN model achieved the highest Recall value (99.00%), its relatively low Precision (90.60%) suggests a tendency to over-predict the majority class, leading to a considerably lower F1-score compared with the proposed approach. This finding highlights the imbalance between Precision and Recall in the TCN model. In contrast, the CNN–BiLSTM–Attention architecture maintains a more favorable trade-off between these metrics, resulting in superior overall classification performance..

4. Conclusions

This study has conducted a comprehensive empirical evaluation and comparative analysis of the performance of three popular deep learning architectures—VGG16, VGG19, and ResNet50—in the automatic classification of eight distinct skin pathologies based on clinical images. To ensure objectivity, all models were trained and evaluated on the same dataset with a unified preprocessing workflow and optimization strategy. The experimental results reveal that VGG16 achieves the highest performance with an accuracy of 93.99%, proving the outstanding effectiveness of extracting complex visual features in medical tasks through homogeneous, small-sized convolutional blocks. The VGG19 model also delivers impressive results with an accuracy of 91.31%, validating the representational capacity of deep-structured network designs. In contrast, ResNet50 yields a significantly lower accuracy of 59.66%, indicating that the performance of networks utilizing skip connections heavily depends on hyperparameter tuning for specific tasks. Moving forward, future research directions will focus on improving the architecture of these deep learning networks to optimize accuracy, aiming for effective practical application within the healthcare system in Vietnam.

References

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.

[2] Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” IEEE Transactions on Biomedical Engineering, vol. 32, no. 3, pp. 230–236, 1985.

[3] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” Journal of Big Data, vol. 6, no. 27, 2019.

[4] D. Yin, et al., “An enhanced hybrid model combining CNN, BiLSTM, and attention mechanism for ECG segment classification,” Frontiers in Physiology, 2024.

[5] D. Yin, et al., “Deep learning-based ECG arrhythmia classification: A systematic review,” Applied Sciences, vol. 13, no. 8, 2023.

[6] A. Vaswani, et al., “Attention is all you need,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.

[7] U. R. Acharya et al., "Application of deep convolutional neural network for automated detection of cardiac health using ECG signals," Computers in Biology and Medicine, vol. 100, pp. 270–278, Sep. 2018

[8] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proc. 32nd Int. Conf. Machine Learning (ICML), Lille, France, 2015, pp. 448–456.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. 3rd Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, 2015. arXiv:1412.6980.

[11] S. Fazeli, “ECG Heartbeat Categorization Dataset,” Kaggle, 2018. Available: https://www.kaggle.com/datasets/shayanfazeli/heartbeat