A METHOD TO IMPROVE THE ACCURACY OF DEEP LEARNING MODELS IN CARDIOVASCULAR DISEASE CLASSIFICATION USING ELECTROCARDIOGRAM SIGNALS

Nguyen Thu Huong, Dinh Cong Tung*, Mai Duc Vinh
University of Transport and Communications

ARTICLE INFO

Received: 29/5/2025
Revised: 30/6/2025
Published: 30/6/2025

KEYWORDS

Electrocardiogram
One-dimensional convolutional neural network
Long short-term memory
Multihead attention
Cardiovascular diseases

ABSTRACT

This paper proposes an improved method to enhance the accuracy of classifying cardiovascular diseases based on electrocardiogram signals by applying a deep learning model composed of multiple integrated components. Specifically, the model architecture is built upon a one-dimensional convolutional neural network to extract local features from raw electrocardiogram signals, effectively capturing significant patterns in the input data. Subsequently, a long short-term memory network is employed to exploit the temporal dependencies within the signal, enabling the model to understand contextual relationships and dynamic changes in features over time. To further improve the model's ability to focus on the most relevant information for classification, a multihead attention mechanism is integrated after the long short-term memory layer. This attention mechanism allows the model to learn the relative importance of different segments within the signal sequence more effectively. Experimental results demonstrate that the combination of one-dimensional convolutional neural network, long short-term memory, and multihead attention yields high performance, achieving an accuracy of over 97% in classifying four types of heart diseases. The proposed method shows promising potential for the application of artificial intelligence in the automated diagnosis of cardiovascular conditions.

MỘT PHƯƠNG PHÁP CẢI THIỆN ĐỘ CHÍNH XÁC CỦA MÔ HÌNH HỌC SÂU TRONG VIỆC PHÂN LOẠI CÁC BỆNH LÝ TIM MẠCH DỰA TRÊN TÍN HIỆU ĐIỆN TÂM ĐỒ

Nguyễn Thu Hường, Đinh Công Tùng*, Mai Đức Vinh
Trường Đại học Giao thông vận tải

THÔNG TIN BÀI BÁO

Ngày nhận bài: 29/5/2025
Ngày hoàn thiện: 30/6/2025
Ngày đăng: 30/6/2025

TỪ KHÓA

Tín hiệu điện tim
Mạng nơ-ron tích chập một chiều
Bộ nhớ dài ngắn hạn
Cơ chế chú ý đa đầu
Bệnh lý tim mạch

TÓM TẮT

Bài báo này đề xuất một phương pháp cải tiến nhằm nâng cao độ chính xác trong việc phân loại các bệnh lý tim mạch dựa trên tín hiệu điện tâm đồ thông qua việc ứng dụng một mô hình học sâu tích hợp nhiều thành phần. Cụ thể, kiến trúc mô hình được xây dựng dựa trên mạng nơ-ron tích chập một chiều để trích xuất các đặc trưng cục bộ từ tín hiệu điện tâm đồ, qua đó nhận diện hiệu quả các mẫu quan trọng trong dữ liệu đầu vào. Tiếp theo, mạng bộ nhớ dài ngắn hạn được sử dụng để khai thác các mối quan hệ theo thời gian trong tín hiệu, giúp mô hình hiểu được ngữ cảnh và sự biến đổi động của các đặc trưng trong quá trình hoạt động của tim. Nhằm nâng cao khả năng tập trung vào những thông tin có liên quan nhất đến nhiệm vụ phân loại, cơ chế chú ý đa đầu được tích hợp sau lớp bộ nhớ dài ngắn hạn. Cơ chế này cho phép mô hình học các trọng số của từng đoạn trong chuỗi tín hiệu một cách hiệu quả hơn. Kết quả thực nghiệm cho thấy sự kết hợp giữa mạng nơ-ron tích chập một chiều, bộ nhớ dài ngắn hạn và cơ chế chú ý đa đầu mang lại hiệu suất cao, đạt độ chính xác trên 97% trong việc phân loại bốn loại bệnh lý tim. Phương pháp được đề xuất cho thấy tiềm năng đầy hứa hẹn trong việc ứng dụng trí tuệ nhân tạo vào hệ thống chẩn đoán tự động các bệnh lý tim mạch.

DOI: https://doi.org/10.34238/tnu-jst.12910

* Corresponding author. *Email: tungdc@utc.edu.vn*

1. Introduction

Cardiovascular diseases are among the leading causes of death worldwide, requiring early and accurate diagnosis to reduce risks and improve treatment effectiveness. Electrocardiography (ECG) is a widely used method for detecting and diagnosing heart diseases. By recording the heart's electrical activity over time, ECG provides essential data that help doctors quickly and accurately assess a patient's cardiac condition. However, ECG signal analysis is still largely manual or semi-automated, relying heavily on the expertise of medical professionals. This dependence can lead to diagnostic errors, especially in cases with complex or noisy signals. Additionally, with the increasing number of cardiovascular patients globally, interpreting large volumes of ECG signals poses a major challenge for healthcare facilities, highlighting the urgent need for fast and accurate automated diagnostic systems. As a result, developing artificial intelligence-based (AI-based) ECG analysis systems with high accuracy and efficiency has become a key research focus in modern medicine.

In recent years, deep learning methods have been widely applied to ECG signal analysis due to their effectiveness in detecting abnormalities. Paper [1] introduces the Artificial Neural Network (ANNet) neural network for detecting cardiac abnormalities from ECG data collected via IoT Edge sensors. The model uses the Synthetic Minority Oversampling Technique (SMOTE) technique to balance training data and combines Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) blocks to predict irregular heartbeats, achieving 97% accuracy with low resource requirements, making it suitable for mobile IoT devices. Paper [2] presents ECG-Attribute-Decomposed Generative Adversarial Network (ADGAN), a model based on Generative Adversarial Network (GAN) combined with bidirectional LSTM to enhance noise reduction and detect abnormal heart rhythms. By using a batch-wise discrimination process, this method preserves variability in anomaly detection, reaching an accuracy of 95.5%. Paper [3] explores transfer learning using convolutional neural networks (CNNs) for ECG signal classification. The authors propose converting signals into image representations such as spectrograms or recurrence plots before feeding them into models like ResNet, Visual Geometry Group (VGG) and Inception for feature extraction. Experimental results show this approach improves accuracy and reduces training time compared to traditional techniques. Paper [4] proposes an unsupervised learning method for heartbeat anomaly detection using transformer layers and dropout mechanisms to prevent overfitting and enhance feature extraction. When tested on ECG5000 and Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) Arrhythmia datasets, the model achieved accuracies of 99% and 89.5%, respectively. Paper [5] introduces a multi-model deep learning system combining CNN-LSTM and RRHOS-LSTM to improve ECG classification while validating results with another CNN-LSTM model to reduce false positives. Experiments on MIT-BIH data showed an accuracy of 95.81%. Paper [6] presents two models, EnsCVDD-Net and BlCVDD-Net, for heart disease detection by integrating LeNet and Gated Recurrent Unit (GRU). EnsCVDD-Net aggregates classification results from LeNet and GRU, while BlCVDD-Net combines these networks as the foundation for an MLP-based ensemble. ECG data is preprocessed with the Adaptive Synthetic Sampling (ADASYN) method to balance sample sizes and Point-Biserial Correlation Coefficient (PBCC) coefficients are used for feature extraction. EnsCVDD-Net achieved 88% accuracy with a processing time of 777 seconds, while BlCVDD-Net reached 91% accuracy in 247 seconds. Paper [7] focuses on early detection of congenital heart disease in children by combining ECG wave data, wavelet features and manually inputted information. The deep learning model, implemented on the Keras platform, achieved a Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score of 0.915 and a specificity of 0.881 when tested on data from the Outpatient Cardiology Department of Guangdong Provincial People's Hospital (China). Paper [8] presents a CNN-based method using Short-Time Fourier Transform (STFT) spectrograms to analyze heart rhythms and detect abnormalities. Two-channel

ECG data from MIT-BIH and European ST-T databases are preprocessed with the Hanning function before being fed into a CNN model. Results showed that ResNet18-Gray achieved 99.79% accuracy, while ResNet34-Gray reached 99.18%. Papers [9]-[11] propose CNN-based models for ECG signal classification, incorporating denoising, normalization, data augmentation and frequency analysis for preprocessing. Experimental results demonstrate that these methods accurately classify heart conditions from ECG data, outperforming traditional machine learning models and showing practical applicability. Paper [12] introduces a denoising autoencoder (DAE) combined with ConvBiLSTM. Using DAE in preprocessing enhances feature prominence by adding and self-removing noise, improving input data quality for ConvBiLSTM. This combination achieved high performance, with a test accuracy of 98%. Paper [13] employed Temporal Convolutional Networks (TCNs) for heartbeat classification on the ECG5000 dataset, achieving an accuracy of 94.2% and improving Balanced Accuracy by 16.5% compared to the state-of-the-art (SoA). Paper [14] applied a Transformer-based architecture with multi-scale shifted windows to extract features from 12-lead ECG signals, achieving a macro F1-score of approximately 77.85% and a sample F1-score of around 81.26% on the PTBXL-2020 dataset. While TCNs offer efficient sequence modeling, they have limitations with fixed receptive fields and limited adaptability to non-uniform EEG patterns. Similarly, Transformers capture long-range dependencies well but are computationally intensive, require large datasets, and risk overfitting due to their high parameter count. Also, both models have problems with noisy, small-scale EEG data and may lack the inductive biases needed for effective temporal feature extraction, making them less ideal in isolation for EEG signal analysis.

Recent studies have demonstrated the strong potential of deep learning models in ECG signal analysis, particularly CNNs for local feature extraction and LSTM networks for modeling temporal dependencies. However, fully capturing long-term dependencies in ECG segments remains challenging, especially in the presence of noise, signal variability, and temporal complexity. To address this, researchers have explored advanced architectures such as Transformer and TCNs. Transformer models, while effective in sequence modeling, typically require substantial computational resources and large-scale annotated datasets to generalize well. Moreover, the self-attention mechanism treats all positions equally, which may lead to overfitting when applied to small or imbalanced datasets. TCNs utilize dilated convolutions to model long-range dependencies more efficiently. However, their reliance on fixed-size convolutional kernels limits their ability to adapt to variable-length dependencies or irregular patterns—characteristics commonly observed in ECG data.

To overcome these limitations, this paper proposes a hybrid model combining one-dimensional convolutional neural networks (1D-CNN), LSTM networks and the multihead attention mechanism for classifying four types of heart conditions. The 1D-CNN extracts local features from ECG signals, enabling the model to recognize key heartbeat patterns. Next, the LSTM learns sequential relationships, enhancing memory retention. Finally, the Multihead Attention mechanism helps the model focus on critical signal regions, improving its ability to differentiate between different arrhythmias, including Sinus Bradycardia (SB), Atrial Fibrillation (AFIB), Sinus Tachycardia (ST) and Normal Sinus Rhythm (SR). Experimental results show that this combination enhances classification accuracy and improves generalization across diverse ECG datasets, contributing to more effective automatic diagnosis in clinical applications.

2. Methods

2.1. One-dimensional convolutional neural networks (1D-CNN)

The 1D-CNN is a deep learning architecture designed for sequential data processing, capable of automatically extracting local features through convolution operations and reducing data dimensionality via pooling. In a 1D-CNN, a convolutional layer slides a filter $\mathrm{W} \in \mathrm{R}^{\mathrm{k}}$ over the input $X \in R^n$, where n is the input length, the output at position i is computed as $y_i = f(\sum_{j=0}^{k-1} W_j X_{i+j} + b)$. Here, k is the kernel size, W_j is the filter weight at index j , $b \in R$ is the bias term and f is the activation function. The Max Pooling layer $y_i = max(X_i, X_{i+1}, \dots, X_{i+k-1})$ reduces data dimensionality while selecting key features. Techniques such as Batch Normalization $\hat{x}_i = \frac{x_i - \mu}{\sigma}$, where μ and σ are the batch mean and standard deviation, and Dropout enhance model stability and reduce overfitting. The general structure of a 1D-CNN consists of multiple convolutional blocks (Conv1D) with activation functions (ReLU), followed by pooling layers and fully connected (FC) layers for classification. The model is optimized using the Cross-Entropy loss function $L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$, where y_i is the ground-truth label, $\hat{y}_i$ is the predicted probability, and N is the number of training samples. The model is trained with the Adam algorithm, which updates weights as $W \leftarrow W - \eta \frac{\partial L}{\partial W}$ where η is the learning rate and $\frac{\partial L}{\partial W}$ is the gradient of the loss with respect to the weights. With its ability to automatically learn features without manual extraction, 1D-CNN is highly effective in classifying sequential data such as ECG signals, audio and time series [15].

2.2. Long short-term memory (LSTM)

LSTM network is a type of recurrent neural network (RNN) designed to handle long-term dependencies in time series data through a gating mechanism, overcoming the vanishing gradient problem in traditional RNNs [16]. At each time step t, the memory state of LSTM is regulated by three key gates: Forget gate $f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$ determines the amount of information to discard from the previous state; input gate $i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$ controls the update of new information into memory and candidate memory state $\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c)$ represents new candidate information to be added; output gate $o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o)$ regulates how much of the memory state contributes to the output. The memory state is updated as $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$ and the hidden output of LSTM is $h_t = o_t \odot \tanh(C_t)$. Here, σ is the sigmoid activation function, $tanh$ is the hyperbolic tangent function, and $\odot$ denotes element-wise multiplication. The input vector $x_t \epsilon R^{d_x}$, the hidden state $h_t \epsilon R^{d_h}$, and the cell state $C_t \epsilon R^{d_h}$ at time t. The matrices W_*, $U_* \epsilon R^{d_h \times d_x}$, and $R^{d_h \times d_h}$, respectively, and the bias vectors $b_* \epsilon R^{d_h}$, where $* \epsilon \{f, i, o, c\}$ are trainable parameters. With this structure, LSTM effectively captures and retains temporal dependencies, enhancing performance in sequence-based tasks such as ECG signal analysis and time series processing.

2.3. Multihead Attention

The Multihead Attention (MHA) mechanism enables models to learn relationships between elements in a sequence by using multiple attention heads in parallel [17]. Given an input sequence $X \in \mathbb{R}^{n \times d_{\text{model}}}$, where n is the sequence length and d_{model} is the model dimensionality, MHA projects the input into queries Q_i, keys K_i, and values V_i for each attention head i via learnable linear projections $Q_i = XW_i^Q$, $K_i = XW_i^K$, $V_i = XW_i^V$, where $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$ are trainable weight matrices for the i^{th} head, and d_k is the dimensionality of the projected space.

Each attention head computes $\text{head}_i = \text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i$. The outputs of all h attention heads are concatenated and projected back to the original dimension using a final linear layer $\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O$, where $W^O \in \mathbb{R}^{h d_v \times d_{\text{model}}}$ is the output projection matrix, d_v is the dimensionality of each value vector, and h is the number of heads. By attending to information from multiple representation subspaces, MHA effectively captures long-range dependencies in sequential data. This makes it suitable for a wide range of applications including natural language processing, computer vision, and biomedical signal analysis such as ECG.

2.4. Proposed method

In this paper, we propose an improved deep learning approach for detecting four types of heart disease by combining 1D-CNN, LSTM and the Multihead Attention mechanism. Figure 1 illustrates the overall architecture of the proposed method. First, the input data is processed through four convolutional blocks, each consisting of two consecutive 1D-CNN layers to extract local features from the signal. After every two convolutional layers, a MaxPooling layer is applied to reduce output size, minimize the number of parameters, prevent overfitting and retain the most significant features. The number of filters in the 1D-CNN layers increases progressively across the blocks (32, 64, 128 and 256) to enhance feature learning capabilities, allowing the model to detect heart rhythm abnormalities more accurately. However, CNN primarily captures local features and struggles to model long-term dependencies in sequential data. To address this limitation, an LSTM layer is integrated after the four convolutional blocks to learn temporal dependencies in ECG signals. Once the features are extracted by LSTM, the Multihead Attention mechanism is applied to enhance the model's focus on crucial signal regions. Not all segments of an ECG signal contain equally valuable information—some segments may include critical pathological markers, while others may contain noise or minor fluctuations. While self-attention mechanisms are known to be effective in capturing global dependencies, a single-head self-attention mechanism may be insufficient to capture the diverse types of relationships present in noisy, non-linear biomedical signals like ECG. In contrast, Multihead Attention improves upon this limitation by employing multiple attention heads in parallel, allowing the model to project the input sequence into different representation subspaces and learn varied attention patterns simultaneously. This parallel mechanism enables the model to capture a richer set of temporal relationships, including subtle, localized features as well as long-range dependencies, which may not be effectively modeled by a single attention head. Furthermore, Multihead Attention has been empirically shown to outperform single-head self-attention and even full Transformer encoders in certain biomedical applications when data is limited or noisy. For example, Kwon et al. [18] demonstrated that Multihead Attention enhances the robustness of EEG signal denoising compared to standard attention mechanisms. Zhao et al. [19] showed that integrating Multihead Attention with CNN-LSTM leads to better generalization in epileptic seizure detection. Similarly, Roy et al. [20] emphasized that multi-representation attention helps focus on clinically relevant regions in physiological signals while suppressing irrelevant fluctuations. Given that the dataset used in this study is relatively small and includes real-world noise, employing a full Transformer encoder could lead to overfitting and instability during training. Therefore, Multihead Attention strikes a balance between model complexity and interpretability, while still significantly improving performance. This design choice not only boosts the model's classification accuracy but also enhances its ability to differentiate between heart rhythms in a clinically meaningful way. Finally, the model's output is passed through a classifier to determine four heartbeat categories: Atrial Fibrillation (AFIB), Sinus Bradycardia (SB), Normal Sinus Rhythm (SR) and Sinus Tachycardia (ST). By integrating 1D-CNN, LSTM and Multihead Attention, the model effectively learns local features, sequential dependencies and optimizes its focus on critical signal segments. This architecture enhances classification accuracy compared to traditional methods.

3. Experimental results

3.1. Dataset

This study utilizes the A 12-lead electrocardiogram database, which contains 12 different cardiac conditions recorded from 10,646 patients, as referenced in [21]. Due to the imbalance in sample distribution across different conditions, the research team focuses on four specific types for model training and evaluation. In details, the dataset includes: SB (1,800 recordings), SR

(1,826 recordings), AFIB (1,780 recordings) and ST (1,568 recordings).

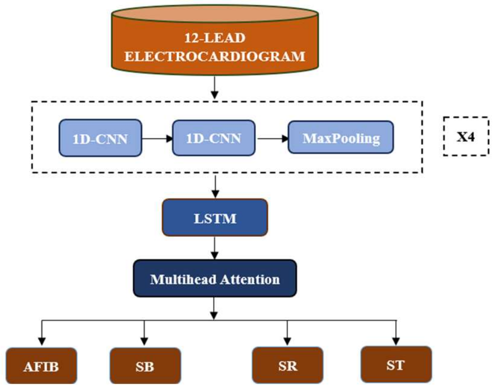


Figure 1. *Overall architecture of the proposed method*

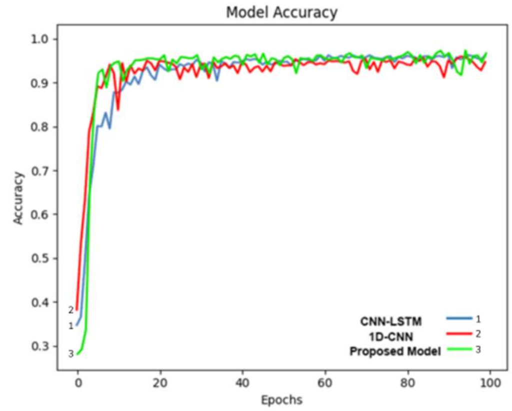


Figure 2. *Accuracy trends of three models with 100 epochs*

3.2. Results

In this section, we present the training and testing results of different models in detecting cardiac conditions. The model training was conducted on a system equipped with an Intel Xeon CPU 2.20GHz, an NVIDIA Tesla T4 GPU with 16GB VRAM, 16GB RAM and Python 3.6. After conducting experiments with various hyperparameter sets, we selected the optimal training configuration as follows: random_state = 42, learning_rate = 1e-4, kernel_size = 5, pool_size = 2, and dropout = 0.2. The Multi-Head Attention module in the proposed model utilizes 8 heads, each with a dimensionality of 8. Before being fed into the model, the data was preprocessed, encoded and normalized. It was then split into training and testing sets in an 80:20 ratio. Figure 2 illustrates the accuracy trends of three models (CNN-LSTM, 1D-CNN and the proposed model) during testing over 100 epochs. The x-axis represents the number of epochs, while the y-axis indicates accuracy. The curves demonstrate how each model's performance evolves over time, allowing for an evaluation of their learning capabilities.

During the initial 0–20 epochs, all models exhibit a rapid increase in accuracy, indicating their ability to learn features effectively from the early stages. The proposed model (green curve) converges faster than CNN-LSTM (1) and 1D-CNN (2), reaching high accuracy within just 10 epochs. In the 20–100 epoch range, all three models achieve high accuracy (> 90%) and stabilize around this value. The 1D-CNN model (2) achieves accuracy close to the proposed model but exhibits greater fluctuations, likely due to its inability to capture long-term dependencies in ECG signals. The CNN-LSTM model (1) converges more slowly than the other two models in the early stages but stabilizes after approximately 40 epochs. This suggests that using LSTM alone, without an attention mechanism, may hinder the model's ability to rapidly learn key features. The proposed model (3) maintains more stable accuracy compared to the other models, with lower fluctuations, demonstrating better generalization on ECG data.

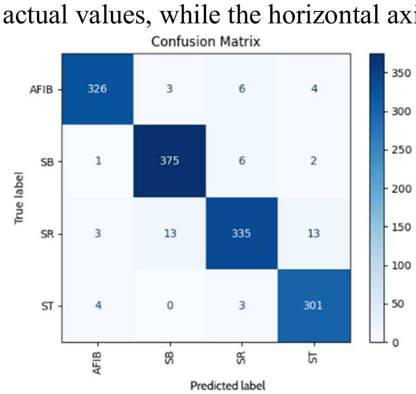
To further demonstrate the effectiveness of the proposed method, we conducted a statistical analysis and compared four evaluation metrics across the trained models: Precision, Recall, F1-Score, and Accuracy. Specifically, Accuracy is defined as the ratio of correctly classified samples to the total number of samples in the test set. Precision measures the proportion of correctly predicted samples for a given class among all samples predicted as belonging to that class. Recall denotes the proportion of correctly predicted samples of a class relative to the total number of actual samples of that class. The F1-Score is the harmonic mean of Precision and Recall, providing a balanced assessment of the model's performance on both metrics [22], [23].

The average values of these four metrics across the target classes are summarized in Table 1.

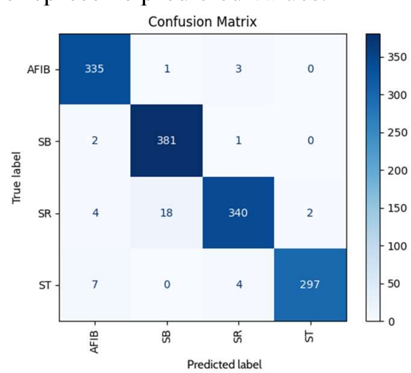
Table 1. *Accuracy comparison results for 3 models*

Model	Accuracy	Precision	Recall	F1-Score
1D-CNN	95.84%	97.60%	97.93%	96.87%
CNN-LSTM	96.98%	96.26%	98.82%	97.52%
Proposed model	97.34%	97.93%	96.16%	97.93%

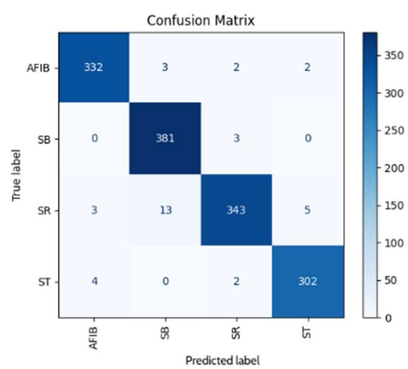
To provide a more intuitive representation of classification accuracy on the test dataset, we utilize confusion matrices in Figures 3a, 3b and 3c, corresponding to the 1D-CNN, CNN-LSTM and the proposed model, respectively. These matrices offer detailed insights into the classification performance of each model for four heartbeat categories: Atrial Fibrillation (AFIB), Sinus Bradycardia (SB), Normal Sinus Rhythm (SR) and Sinus Tachycardia (ST). The vertical axis represents actual values, while the horizontal axis represents predicted values.



(a)



(b)



(c)

Figure 3. *Confusion matrix of (a) 1D-CNN model, (b) CNN-LSTM model, (c) proposed model*

The confusion matrices indicate that the 1D-CNN model performs well overall, especially in identifying the SB class, but struggles with distinguishing SR from other classes. The CNN-LSTM model improves classification accuracy, particularly for AFIB and SB, though misclassification between SR and SB remains an issue. The proposed model, which combines 1D-CNN, LSTM, and Multihead Attention, achieves the best overall performance, significantly reducing misclassification across all classes - especially for AFIB and SR - demonstrating its superior capability in accurately identifying different heart rhythms.

To provide a more objective and comprehensive evaluation of the performance of each model, we conducted a statistical analysis using one-way ANOVA to compare the classification performance among the three models. Specifically, each model (1D-CNN, CNN-LSTM, and the proposed model) was trained and evaluated five times to ensure the stability of the results. The ANOVA test yielded a p-value of 0.0122, which is lower than the significance level of 0.05. This indicates that the performance differences among the models are statistically significant and not due to random variation. These findings further support the effectiveness and the well-founded design choice of the integrated model proposed in this study.

4. Conclusions

This paper proposes a method that combines 1D-CNN, LSTM and the Multihead Attention mechanism to classify cardiac arrhythmias using data from a 12-lead electrocardiogram. The model focuses on identifying four heart conditions: Sinus Bradycardia (SB), Atrial Fibrillation (AFIB), Sinus Tachycardia (ST) and Normal Sinus Rhythm (SR). In this approach, CNN layers are used to extract features, LSTM captures temporal dependencies and Multihead Attention enhances the model's focus on critical signal segments, thereby improving classification accuracy. Experimental results show that the proposed method achieves an accuracy of 97.34%, outperforming previous models and highlighting the potential of deep learning in automatically detecting heart diseases from ECG signals. In the future, we aim to further optimize the model and enhance its accuracy to better support medical diagnosis and treatment.

Acknowledgements

This research is funded by University of Transport and Communications (UTC) under grant number T2025-CN-002.

REFERENCES

[1] G. Sivapalan, K. K. Nundy, S. Dev, B. Cardiff, and D. John, "ANNet: A lightweight neural network for ECG anomaly detection in IoT edge sensors," *IEEE Trans. Biomed. Circuits Syst*., vol. 16, no. 1, pp. 24–35, Jan. 2022, doi: 10.1109/TBCAS.2021.3137646.

[2] J. Qin, F. Gao, Z. Wang, D. C. Wong, Z. Zhao, S. D. Relton, and H. Fang, "A novel temporal generative adversarial network for electrocardiography anomaly detection," *Artif. Intell. Med.,* vol. 136, Jan. 2023, Art. no. 102489, doi: 10.1016/j.artmed.2023.102489.

[3] L. Mohebbanaaz, V. R. Kumar, and Y. P. Sai, "A new transfer learning approach to detect cardiac arrhythmia from ECG," *Signal Image Video Process.,* vol. 16, no. 7, pp. 1945–1953, Jul. 2022, doi: 10.1007/s11760-022-02155-w.

[4] A. Alamr and A. Artoli, "Unsupervised transformer-based anomaly detection in ECG signals," *Algorithms,* vol. 16, no. 3, Mar. 2023, Art. no. 152, doi: 10.3390/a16030152.

[5] X. Zhang, Z. Huo, and Q. Wu, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," *IEEE Access*, vol. 9, pp. 101746–101759, Jul. 2021, doi: 10.1109/ACCESS.2021.3096610.

[6] H. Khan, N. Javaid, T. Bashir, M. Akbar, N. Alrajeh, and S. Aslam, "Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: EnsCVDD-Net and BlCVDD-Net," *IEEE Access*, vol. 12, pp. 109230–109254, Jul. 2024, doi: 10.1109/ACCESS.2024.3421241.

[7] J. Chen, S. Huang, Y. Zhang, *et al*., "Congenital heart disease detection by pediatric electrocardiogram based deep learning integrated with human concepts," *Nat. Commun*., vol. 15, Jan. 2024, doi: 10.1038/s41467-024-44930-y.

[8] H. Li and P. Boulanger, "Structural anomalies detection from electrocardiogram (ECG) with spectrogram and handcrafted features," *Sensors,* vol. 22, no. 7, Mar. 2022, Art. no. 2467, doi: 10.3390/s22072467.

[9] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.,* vol. 405, pp. 81–90, Sep. 2017, doi: 10.1016/j.ins.2017.04.012.

[10] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint*, arXiv:1707.01836, Jul. 2017.

[11] O. Yildirim, U. B. Baloglu, R. S. Tan, and U. R. Acharya, "A deep convolutional neural network

model for automated identification of abnormal heart rhythms using ECG signals," *Appl. Soft Comput*., vol. 84, Oct. 2019, Art. no. 105619, doi: 10.1016/j.asoc.2019.105619.

[12] B. Tutuko, A. Darmawahyuni, S. Nurmaini, A. E. Tondas, M. N. Rachmatullah, *et al*., "DAE-ConvBiLSTM: End-to-end learning single-lead electrocardiogram signal for heart abnormalities detection," *PLoS ONE*, vol. 17, no. 12, Dec. 2022, Art. no. e0277932, doi: 10.1371/journal.pone.0277932.

[13] T. M. Ingolfsson, H. A. A. Madsen, A. Laursen, and P. Popovski, "ECG-TCN: Wearable Cardiac Arrhythmia Detection with a Temporal Convolutional Network," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 17, no. 2, pp. 225–238, Apr. 2023, doi: 10.1109/TBCAS.2023.3240431.

[14] R. Cheng, S. Li, Z. Zhang, J. Chen, and X. Hu, "MSW-Transformer: Multi-Scale Shifted Windows Transformer Networks for 12-Lead ECG Classification," *arXiv preprint* arXiv:2311.13583, 2023. [Online]. Available: https://arxiv.org/abs/2311.13583. [Accessed Feb. 15, 2025].

[15] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in *Proc. 2017 IEEE 19th Int. Conf. e-Health Netw., Appl. Serv. (Healthcom)*, Dalian, China, Oct. 2017, pp. 1–6, doi: 10.1109/HealthCom.2017.8210842.

[16] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci. Rep.,* vol. 11, Mar. 2021, doi: 10.1038/s41598-021-86390-1.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.,* vol. 30, pp. 5998–6008, Dec. 2017.

[18] T. Kwon, H. Lee, J. Kim, and B. Lee, "Transformer-based stacked multi-head attention model for EEG signal denoising," *Brain Informatics*, vol. 10, no. 1, pp. 1–13, 2023.

[19] S. Roy, A. Kiral-Kornek, and D. M. Harrer, "Deep learning enabling technology for epileptic seizure detection using EEG," in *Proc. IEEE EMBC*, 2019, pp. 1–4.

[20] Z. Zhao, Q. Zhang, H. Liu, L. Peng, and Y. Li, "Epileptic Seizure Detection Based on Multi-Head Attention Mechanism and CNN-LSTM Network," *IEEE Access*, vol. 10, pp. 19424–19435, 2022.

[21] J. Zheng, J. Zhang, S. Danioko, *et al.,* "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Sci. Data,* vol. 7, Feb. 2020, doi: 10.1038/s41597-020-0386-x.

[22] C. W. Cleverdon, "Factors Determining the Performance of Indexing Systems," Aslib Cranfield Research Project, 1966.

[23] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Butterworth-Heinemann, 1979.