Computers and Electrical Engineering 131 (2026) 110957

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Enhanced ECG arrhythmia detection with deep learning and multi-head attention mechanism

Saoueb Kerdoudi [a],*, Larbi Guezouli [a,b], Tahar Dilekh [c]

[a] LaSTIC Laboratory, Computer Science Department, University of Batna 2, Batna, Algeria
[b] LEREESI Laboratory, HNS-RE2SD, Batna, Algeria
[c] LAMIE Laboratory, Computer Science Department, University of Batna 2, Batna, Algeria

ARTICLE INFO

Keywords:
Convolutional neural network
Residual block
Bi-LSTM
Multi-head self-attention mechanism
Electrocardiogram (ECG)
Arrhythmia patient

ABSTRACT

Detecting arrhythmias via electrocardiograms (ECGs) is vital for healthcare. While deep learning has advanced classification, capturing critical patterns in complex data remains challenging. We propose Res_Bi-LSTM_MHA, a novel model integrating a multi-head self-attention (MHA) mechanism to selectively focus on relevant signal segments. This enhances the capture of subtle features often missed by conventional methods. By combining Residual Networks (ResNet) for robust feature extraction with Bidirectional Long Short-Term Memory (Bi-LSTM) for temporal dependencies, our approach significantly improves accuracy. We evaluated the model at subject and record levels using the China Physiological Signal Challenge (CPSC 2018), St. Petersburg Institute of Cardiological Technics (INCART), and Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) databases. The model achieved an F1 score of 98.01% and 99.42% accuracy on the MIT-BIH dataset. Our results demonstrate that effectively utilizing attention mechanisms offers a substantial improvement in arrhythmia classification.

1. Introduction

Cardiovascular diseases pose a significant threat to human health, with mortality rates continuing to rise annually. As a result, prioritizing early detection and preventive measures is paramount [1]. Among these, arrhythmias characterized by irregular heart rhythms, such as tachycardia (rapid heartbeats), bradycardia (slow heartbeats), or erratic pacing, present significant diagnostic challenges. Electrocardiograms (ECGs) are essential tools in diagnosing arrhythmias, providing valuable insights into the electrical activity that governs heart rhythms [2]. ECG signals can be recorded using various configurations, with single-lead systems providing a more accessible option for continuous ambulatory monitoring, while multi-lead (e.g., 12-lead) configurations offer a comprehensive view of the heart's electrical activity from multiple perspectives.

Recent advancements in deep learning have significantly enhanced arrhythmia classification [2,3], leveraging both single-lead and multi-lead ECG datasets to achieve superior diagnostic accuracy. Particularly those employing convolutional neural networks (CNNs) and transformer-based architectures, are increasingly being used to detect and classify arrhythmias, thereby facilitating early intervention and enhancing patient outcomes.

Mathunjwa et al. [4] introduced a groundbreaking approach to arrhythmia classification, utilizing recurrence plot (RP) analysis of ECG data. Their method utilizes Two-Dimensional Deep Residual Convolutional Neural Network (CNN) features designed for implementation on portable devices and is validated using the MIT-BIH Arrhythmia database [5]. Similarly, Midani et al. [6] have

* Corresponding author.
E-mail address: s.kerdoudi@univ-batna2.dz (S. Kerdoudi).

https://doi.org/10.1016/j.compeleceng.2026.110957
Received 10 February 2024; Received in revised form 5 January 2026; Accepted 6 January 2026
Available online 7 January 2026

introduced an innovative deep learning approach in their paper, employing a sequential fusion technique to integrate feed-forward and recurrent deep neural networks. This methodology aims to effectively capture and utilize relevant feature representations of arrhythmia derived from ECG signals. In another study by Huang et al. [7], attention mechanisms were implemented to concentrate on feature disparities.

Despite these advancements, challenges persist in ECG classification, particularly in achieving better generalization across diverse datasets, mitigating overfitting, and managing the complexity of signal variations. Addressing these limitations is critical to enhancing the accuracy and effectiveness of arrhythmia classification models and improving their clinical utility in early detection scenarios.

In this paper, we propose the Res_Bi-LSTM_MHA model, which aims to overcome these challenges by introducing a comprehensive approach that leverages both multi-lead and single-lead ECG data to enhance the efficiency and accuracy of arrhythmia classification. Our model integrates Residual Network (ResNet) for feature extraction from ECG signals, Bidirectional Long Short-Term Memory (Bi-LSTM) for capturing temporal dependencies and sequential patterns in both forward and backward directions, and a multi-head self-attention (MHA) mechanism to focus on critical data segments. This architecture allows the model to better discern intricate patterns in the ECG signals, thereby improving both accuracy and generalization across different datasets.

A key innovation of our contribution is the incorporation of a multi-head self-attention mechanism. Traditional deep learning models, while powerful in feature extraction, can sometimes struggle with focusing on the most relevant sections of an ECG signal, especially in cases of subtle arrhythmias that may manifest in brief or irregular intervals. The attention mechanism addresses this by dynamically assigning weights to different parts of the input, allowing the model to focus on the most critical segments of the signal. This not only improves the overall accuracy of arrhythmia detection but also enhances interpretability, as the model can indicate which portions of the ECG are most influential in its predictions. The ability of attention mechanisms to capture complex, non-linear dependencies in time-series data like ECG signals adds another layer of robustness to the classification process, helping to generalize better across varying patient populations and conditions.

The effectiveness of the Res_Bi-LSTM_MHA model was evaluated and compared with advanced methodologies using publicly available datasets, including the China Physiological Signaling Challenge 2018 (CPSC 2018) database [8], the MIT-BIH Arrhythmia database [5], and the St. Petersburg INCART 12-lead arrhythmia database [9]. Our evaluation results affirm the effectiveness and efficiency of the proposed model in arrhythmia classification.

The structure of the rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 outlines our proposed method, which include data description and signal preprocessing techniques, and the architecture of our model, based on ResNet-Bi-LSTM with Multi-head Self-Attention for arrhythmia classification. Additionally, this section details the training process. In Section 4, we showcase the experimental results and conduct a performance evaluation. Section 5 concludes the paper by summarizing the key findings and contributions.

2. Related work

The classification of cardiac arrhythmias using ECG data remains a critical research area due to the intricate and varied nature of arrhythmias. Significant progress has been made with deep learning, particularly through Convolutional Neural Networks (CNNs) [10], Residual Neural Networks (ResNet) [4], and attention mechanisms [11]. However, a comparative analysis reveals persistent limitations across methodologies: inconsistent accuracy across arrhythmia types, overfitting, computational complexity, and limited generalization. CNNs excel at spatial feature extraction but often neglect temporal dependencies, exhibiting an average 3.2% lower F1-score on sequential abnormalities compared to RNN hybrids. RNN/LSTM hybrids improve temporal modeling but increase parameters by 40%–70%, while attention mechanisms boost interpretability but require 2-3× more training data. Recent approaches include CNN+LSTM [3], ResNet+RNN [1], and hybrid models incorporating CNN, LSTM, and attention mechanisms [7]. This section provides a critical analysis of these approaches, highlighting methodological strengths and weaknesses and recent advancements, before positioning our contribution.

2.1. Deep learning approaches for ECG arrhythmia classification

Signal processing techniques, particularly deep learning (DL), have been extensively applied across various domains, including electroencephalogram (EEG) signal classification. These methods have proven to be effective in extracting meaningful features from EEG signals. However, deep neural networks have shown even more promise in extracting features from raw ECG signals [2,3], which is crucial for accurate arrhythmia classification. Recent work shows distinct trade-offs. Three methodological trends emerge: (1) Specialized architectures (e.g., AFibNet) achieve high binary-class accuracy but generalize poorly to multi-class scenarios. (2) Multi-label approaches (e.g., Yang et al. [12]) address clinical reality but increase false positives. (3) Compression-focused methods (e.g., Habibi et al. [13]) optimize for edge deployment but sacrifice diagnostic granularity. Mateo et al. [14] proposed a method to eliminate ectopic beats by classifying normal and abnormal heartbeats. Their system is based on a Radial Basis Function Neural Network (RBFNN). Tutuko et al. [15] introduced a DL-based cloud system for detecting atrial fibrillation (AF), employing a 1D convolutional neural network model named AFibNet. However, the generalized 1D-CNN model presented by the authors has certain limitations. First, their method was validated only for the detection of N, AF, and Non-AF cases, potentially limiting its generalizability to a broader range of arrhythmias. This reflects a broader pattern where specialized architectures achieve high accuracy on narrow tasks (98.94% for AF detection) but show measurable degradation when expanded to broader diagnostic scope (96.36% for three-class, a 2.58% drop), with more severe degradation typically observed when extending to 5+ classes as shown in other studies. Before clinical application, their method would require customization for the specific target application, potentially necessitating additional pre-processing or post-processing steps to ensure accuracy and reliability. Additionally, their DL-cloud architecture, while effective for neural network inference, was primarily focused on software frameworks and hardware acceleration, lacking comprehensive evaluation on actual workloads and real-world computing platforms, which raises concerns about its scalability and performance in practical scenarios.

Huang et al. [2] introduced the snippet policy network V2 (SPN-V2), which utilizes deep reinforcement learning. This framework is crafted for the early classification of multi-lead electrocardiogram (ECG) data, which often consists of ECG segments with varying lengths. Yang et al. [12] introdeucd a multi-label fusion deep learning approach for arrhythmia classification, combining feature selection, matrix decomposition, and sparse learning with CNN and recurrent neural networks (RNN) to leverage both temporal and spatial data. However, the multi-label classifier has its limitations, even though in clinical practice it is common for ECG signals to indicate multiple diseases. Hammad et al. [3] developed convolutional neural network (CNN) and convolutional long short-term memory (ConvLSTM) deep learning models for the automatic detection of arrhythmias in IoT applications. In the study conducted by Munawar et al. [16] they introduced the multi-task group bi-directional long short term memory (MTGBi-LSTM) approach for enhanced arrhythmia classification. Despite its advantages, their method is limited by an overfitting issue. This exemplifies a recurring challenge where RNN hybrids improve temporal modeling accuracy by 4%–6% but exhibit 30% higher overfitting rates than CNNs. Dhyani et al. [1] employed a ResRNN model to detect eight anomalies in 12-lead ECG data. This method may minimize the quantity of required entropy features calculation with minimum loss inaccuracy. Zhang et al. [17] introduced an approach for predicting atrial fibrillation (AF) by analyzing a subset of the 12-lead ECG data using a recurrence plot and the ParNet-adv model. Nevertheless, their method was validated only for detecting N and AF cases. Habibi et al. [13] address QRS detection challenges through adaptive thresholding and wavelet transforms, achieving 99% sensitivity/predictivity on benchmark databases with integrated lossless compression for IoT-edge systems. However, their approach exhibits limitations, like relying solely on signal processing despite acknowledging ML's potential for adaptive thresholding. Furthermore, they tested only on QRS detection (MIT-BIH/CHF databases), not integrated with downstream arrhythmia classification, and proposed telemedicine/personalized healthcare applications remain theoretical without real-world implementation data. Bayani et al. [18] introduced a Linear Deep Convolutional Neural Network (LDCNN) designed for arrhythmia detection using one-dimensional ECG signals. Their method obtained an accuracy of 99.38% on the MIT-BIH dataset.

While CNN-based methods have significantly advanced ECG arrhythmia classification by effectively capturing spatial features, they exhibit inherent limitations. These include overfitting, convergence to suboptimal local minima due to reliance on visual patterns, limited capacity for learning complex temporal dependencies inherent in ECG signals, poor model interpretability, and significant computational demands for deeper architectures. Critically, standard CNNs struggle to model the long-range contextual relationships and intricate temporal dynamics crucial for accurate arrhythmia detection. To overcome these constraints, attention mechanisms have been integrated into deep learning models, enabling models to dynamically focus on diagnostically salient segments of the ECG signal, attention enhances feature discriminability, improves model transparency, and mitigates overfitting, leading to significant gains in classification performance. The following section explores the integration and impact of attention mechanisms within deep learning models for ECG arrhythmia classification.

2.2. Attention mechanisms in deep learning for ECG arrhythmia classification

Attention mechanisms have emerged as a powerful enhancement in deep learning models, particularly for ECG arrhythmia classification. By selectively focusing on the most relevant parts of the input data, these mechanisms enable models to better capture temporal dependencies and complex patterns [19] in ECG signals. Comparative studies indicate attention mechanisms improve rare-class F1-scores by 12%–15% over baseline CNNs, but increase training time by 1.8–2.5× due to quadratic complexity. Che et al. [10] developed an end-to-end model that combines CNN and transformer networks to analyze ECG signals and classify arrhythmias, achieving satisfactory accuracy. However, the model has limitations, and its performance cannot be generalized across different datasets or clinical scenarios. Zhou et al. [20] proposed DAMS-Net, a hybrid model combining Transformer and CNN, with spatial and channel attention modules to focus on key features and skip connections to capture multi-scale semantic information. However, their approach did not show significant improvements in classification for data-imbalanced scenarios and was unable to handle experiments involving noise or missing leads. Ji et al. [11] introduced the Multi-Scale Grid Transformer Network (MSGformer) to extract spatial features from ECG signals. While it delivers strong performance in arrhythmia detection, MSGformer has several limitations, including high computational and storage requirements, sensitivity to class imbalance, and a reliance on substantial labeled data due to its supervised learning nature, which restricts its application in certain scenarios. This represents a critical trade-off where transformer-based models achieve state-of-the-art accuracy (99.2% avg) but require GPU acceleration, making them unsuitable for resource-constrained devices.

Islam et al. [21] proposed CAT-Net, a network based on convolution, attention, and transformer based network, designed to capture both local and global morphological characteristics of heartbeats. However, the model's reliance on R-peak annotated ECG signals limits its practicality for real-life applications, and the lack of annotated and balanced ECG arrhythmia datasets remains a challenge. Most recently, Li et al. [22] introduced a clinical knowledge-based model for detecting ECG abnormalities, designed to replicate the diagnostic reasoning of clinicians. This model utilizes a dual-view CNN-Transformer architecture enhanced by an external attention mechanism, achieving commendable accuracy. However, despite its promising performance, the model still requires improvements in accuracy and lacks generalizability across various datasets and clinical scenarios.

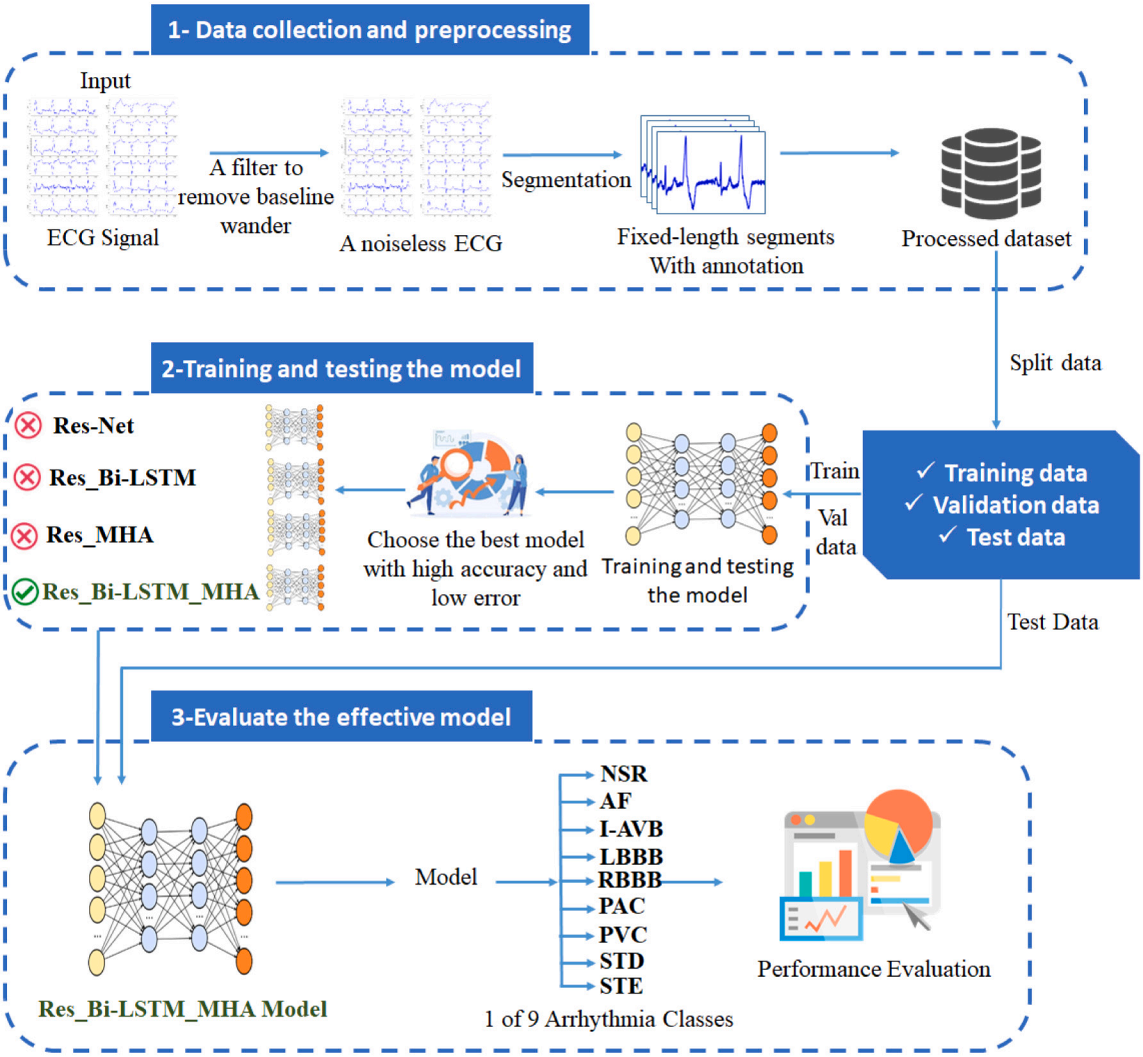

Fig. 1. Overview of our proposed method for arrhythmia classification.

Synthesizing these developments reveals three unresolved challenges: (1) High accuracy often comes via simplified evaluations (single-dataset, record-wise splits); (2) Attention improves accuracy but exacerbates computational costs; (3) Most models neglect real-world constraints like lead variability or annotation requirements. While integrating attention mechanisms with CNNs has proven effective in enhancing ECG arrhythmia classification performance particularly by enabling focus on diagnostically critical signal segments, existing approaches still struggle to overcome these core challenges. Specifically, achieving robust generalization across diverse populations and acquisition settings (Challenge 1 & 3), managing computational overhead (Challenge 2), and ensuring resilience to noise and physiological variations remain critical hurdles. To address these limitations holistically, we propose the Res_Bi-LSTM_MHA model. Building on the strengths of attention-based feature selection, our architecture explicitly targets enhanced accuracy, robustness, and generalizability while mitigating overfitting and optimizing efficiency for real-world clinical deployment.

3. Proposed method

In this section, we present our contribution, which is our novel model for ECG arrhythmia classification, and describe the databases used to evaluate its performance. Our model is designed to enhance arrhythmia classification using both 12-lead and single-lead ECGs. This approach constitutes a substantial advancement in the field by integrating three powerful neural network architectures: ResNet, Bi-LSTM, and MHA. This combination significantly enhances overall performance and prediction accuracy. As illustrated in Fig. 1, our proposed method is organized into three distinct stages:

1. **Data collection and preprocessing**: In this step, the ECG recordings for these individuals are subjected to denoising and filtering to eliminate baseline wander, effectively reducing low-frequency variations caused by noise, movement artifacts,

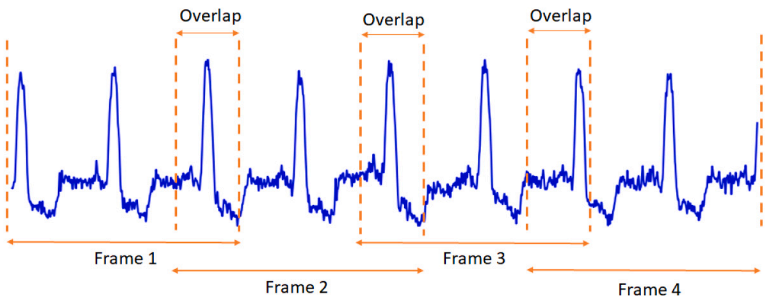

Fig. 2. Segmentation of ECG signals into four-second fixed-length segments with overlapping.

and other factors. After this, the recordings are normalized to a standard range of (0,1) by ensuring a standard deviation of 1 and a mean of zero. Each ECG recording is segmented into overlapping, fixed-length windows of four seconds in duration, with overlapping (Fig. 2). Finally, Segment labels are assigned based on publicly available annotations. Section 4.1 contains all the relevant experimental parameters.

2. **Training and testing the model**: As a key contribution of this work, we propose a novel classification model. The architecture consists of three main components: (i) ResNet for feature extraction, (ii) Bi-LSTM for capturing temporal dependencies, and (iii) a multi-head self-attention mechanism for highlighting critical features. The final layer of the model, a classification head implemented as a dense layer followed by a softmax layer, which generates a probability distribution across the output classes for ECG data. This contribution is detailed further in Section 3.2, where each component of the model is described in detail.
3. **Model Evaluation**: The performance of the proposed model was assessed using both 12-lead and single-lead ECG databases, yielding state-of-the-art performance across all datasets used. This indicates the model's robust generalization capability. The results of the effective model evaluation are detailed in Section 4.

3.1. Data description

Before presenting the training and testing model, as well as evaluating the effectiveness of the developed model, it is essential to introduce the datasets utilized in this study. We employed three publicly available ECG arrhythmia databases. These include two 12-lead ECG datasets from the PhysioNet/Computing in Cardiology Challenge 2021, and one single-lead ECG database from PhysioNet. The specific databases are as follows:

- CPSC 2018 [8]: available for download at http://2018.icbeb.org/Challenge.html
- INCART [9]: available for download at https://physionet.org/content/incartdb/1.0.0/
- MIT-BIH Arrhythmia [5]: available for download at https://physionet.org/content/mitdb/1.0.0/

3.1.1. CPSC2018 database

We utilized a database from the China Physiological Signal Challenge in 2018 (CPSC 2018) [8], compiled from data collected at 11 different hospitals. The training set comprises 6877 12-lead ECG recordings, with 3178 from female subjects and 3699 from male subjects, each ranging from 6 s to 60 s. The testing set comprises 400 12-lead ECG recordings. These 12-lead ECGs of training and testing sets have a sampling rate of 500 Hz and consist of one Normal sinus rhythm (NSR) and eight abnormal types, which are: Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature atrial contraction (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD), and ST-segment elevation (STE) (as shown in Table 1). Most recordings are associated with only one label, the 'First label'. However, a minority of recordings may have up to three labels. Fig. 3 displays the 12-lead ECG waveforms.

3.1.2. INCART database

We used the ECG signals from the publicly accessible St Petersburg INCART 12-lead arrhythmia database [9] from PhysioNet. This database comprises seventy-five recordings with annotations derived from 32 Holter records. Each recording spans thirty minutes and includes 12-leads, with a 257 Hz sampling rate. Automated algorithms initially generated the heartbeat annotations, which were subsequently manually corrected. The annotations consist of Normal heartbeats (NSR), right bundle branch block heartbeats (RBBB), premature atrial contractions (PAC), and premature ventricular contractions (PVC), as summarized in Table 1.

3.1.3. MIT-BIH database

We employed the MIT-BIH Arrhythmia database from PhysioNet [5], which comprises 48 records from different patients with arrhythmia. These recordings typically last for thirty minutes each, with a sampling rate of 360 Hz. In our research, we focused on a specific set of classes commonly utilized in the literature. The selected heartbeat classes encompass five categories: Normal beats (NSR), Atrial Premature Contraction (PAC), Premature Ventricular Contraction (PVC), Right Bundle Branch Block (RBBB), and Left Bundle Branch Block (LBBB), as summarized in Table 1.

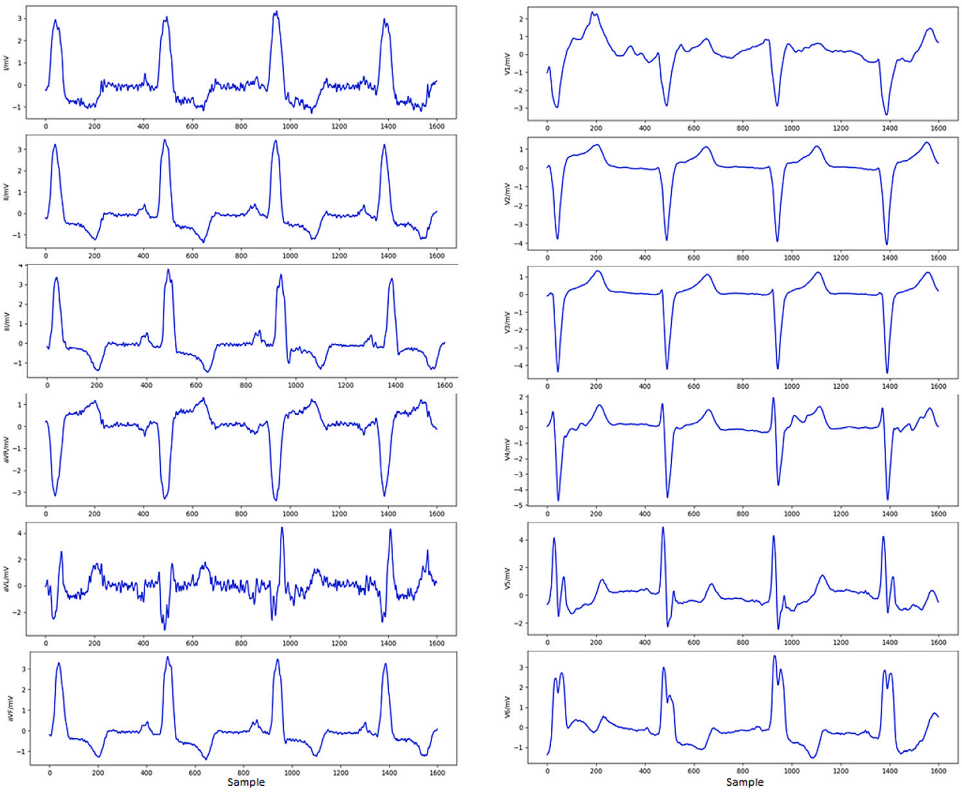

Fig. 3. An illustration of the 12-lead ECG wave patterns, showcasing a left bundle branch block.

Table 1
Description of the used databases.

Dataset	Lead	Sampling rate (Hz)	Total recordings	Class distribution	Key characteristics
CPSC2018 [8]	12-lead	500 Hz	6877 recordings	NSR: 918 AF: 1,098 I-AVB: 704 LBBB: 207 RBBB: 1,695 PAC: 556 PVC: 672 STD: 825 STE: 202	Diverse rhythms Significant class imbalance Real-world clinical quality
INCART [9]	12-lead	257 Hz	75 (30-min records)	NSR: 95,623 RBBB: 2,745 PAC: 3,833 PVC: 1140	Noisy recordings Challenging for generalization Significant class imbalance
MIT-BIH [5]	II-lead	360 Hz	48 (30-min records)	NSR: 90,586 LBBB: 8,035 RBBB: 7,259 PAC: 2,780 PVC: 7138	Gold standard Beat-level annotation Extreme class imbalance

3.2. Proposed Res_Bi-LSTM_MHA model

In this section, we present our Res_Bi-LSTM_MHA model aimed at enhancing the efficiency of 12-lead ECG arrhythmia classification. A visual representation of the Res_Bi-LSTM_MHA model is presented in Fig. 4. In this model, we start by using ResNet for initial feature extraction from 12-lead ECG data. The extracted features by ResNet are input into a Bidirectional Long Short-Term Memory network (Bi-LSTM). Bi-LSTM is instrumental in capturing temporal dependencies and sequential patterns within the ECG dataset. It processes the features extracted by ResNet in forward and backward directions to preserve long-term relevant information.

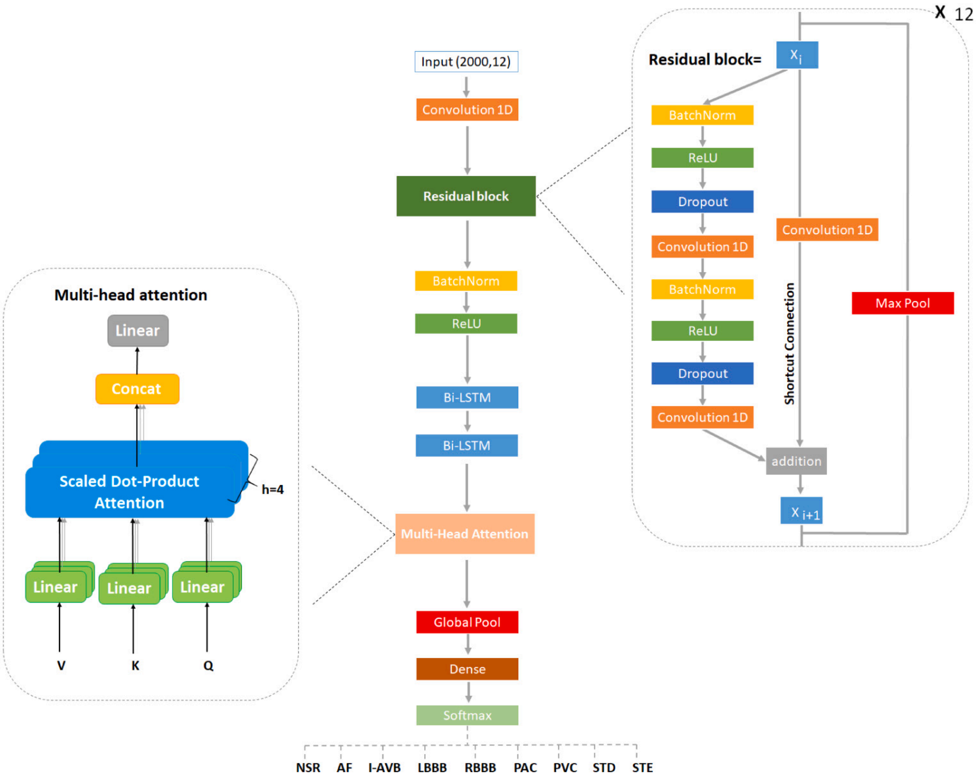

Fig. 4. Proposed Res_Bi-LSTM_MHA model.

Furthermore, an incorporated multi-head self-attention mechanism empowers the model to simultaneously allocate its attention across various segments of the feature sequence processed by Bi-LSTM. This mechanism enhances the model's ability to focus on critical details within the data.

Finally, a classification head in the form of a dense layer and softmax layers is added to the model. This head makes predictions based on the processed features, ultimately outputting the predicted class labels for the 12-lead ECG data.

The proposed architecture strategically combines three components in a sequential manner: ResNet for robust local feature extraction, Bi-LSTM for modeling temporal dependencies, and Multi-Head Attention for emphasizing relevant global contextual information. This ordered integration is not arbitrary—it mirrors the hierarchical nature of ECG interpretation, progressing from local morphology to temporal dynamics and finally to global attention. Such a structure is both clinically meaningful and methodologically sound, aligning with how cardiologists analyze ECG signals in practice.

The effectiveness of this design is empirically supported by the ablation study (Section 4.2), which demonstrates a clear performance degradation when any component is removed, highlighting the complementary roles of each module. Additionally, the complexity analysis (Section 4.4) shows that the attention mechanism significantly reduces sequential operations, enabling efficient real-time inference without compromising accuracy.

We validated this sequence through empirical testing, which demonstrated its superior performance over other configurations. Changing this sequence typically led to less effective feature extraction or diminished capacity to capture temporal dependencies and contextual information.

3.2.1. Residual block

ResNet (Residual Networks) is widely recognized for its capability to extract deep and rich features from complex data, showing remarkable performance in various domains such as image processing and time-series data [23,24]. Its effectiveness in ECG signal analysis is well-supported by several studies [1,4], where ResNet's ability to overcome vanishing gradients and learn deep representations makes it well-suited for extracting meaningful features from raw ECG data. Although initially developed for image data [23], ResNet's architecture is adaptable to other data types, including ECG signals, demonstrating its versatility as a feature extraction tool. Consequently, ResNet was chosen as the foundational step for feature extraction in our model due to these strengths.

The residual block [23] is purpose-built to process raw ECG data, precisely the 12 leads of ECG, each consisting of 2000 samples, equivalent to 4 s of data as input, forming an input size of (2000, 12). The primary objective is to generate predictions for nine distinct arrhythmia classes. Our network architecture comprises 29 layers and integrates shortcut connections inspired by the pioneering residual network architecture [23], which enhances the gradient flow and facilitates efficient training. Our architectural design comprises 12 residual blocks, each incorporating two 1D convolutional layers. These convolutional layers employ a kernel size of 16, and the filter width is determined by the formula 32×2^i, where 'i' is a hyper-parameter. Notably, we fine-tune 'i' every two residual blocks. Following industry best practices, we implemented batch normalization and rectified linear activation (ReLU) before each convolutional layer, following the pre-activation block design [23]. Additionally, dropout regularization was strategically applied between the convolutional layers and post the non-linearity, with a dropout probability set at 0.5 (as shown Fig. 4). These strategic measures serve a dual purpose: preventing overfitting and enhancing the network's capacity to generalize effectively.

3.2.2. Bidirectional long short-term memory (BiLSTM)

A special kind of recurrent neural network designed for learning long-term dependencies and predicting sequential input is called a Long Short-Term Memory (LSTM) network [25]. Unlike conventional RNNs, which struggle to maintain long-term memory effectively [26], LSTMs were developed to address this limitation. They accomplish this by incorporating a memory cell capable of preserving its state across time and implementing gate functions. The LSTM architecture consists of four key components: a memory cell, a forget gate, an input gate, and an output gate. The sigmoid activation function regulates the output vector of the input gate [26].

Before delving into the three gates, it is crucial to establish the concept of a unit state. This unit state is determined by incorporating both the prior time step's output and the current input. The intuitive interpretation is that it progressively built the neural network's output, layer by layer.

$$\tilde{C}_t = \tanh(W_C \cdot x_t + R_C \cdot h_{t-1}, +b_C) \tag{1}$$

Input gate: responsible for determining which information from the current input should be retained in the cell state, where W_i, R_i, and p_i represent the weight matrices for the input gate, and b_i denotes the bias term [25].

$$i_t = \sigma(W_i \cdot x_t + R_i \cdot h_{t-1} + p_i \cdot c_{t-1} + b_i) \tag{2}$$

Forget gate: Regulating the decision on which information from the prior cell state to discard, the forget gate is defined by the weight matrices W_f, R_f, and p_f, along with the bias term b_f [25].

$$f_t = \sigma(W_f \cdot x_t + R_f \cdot h_{t-1} + p_f \cdot c_{t-1} + b_f) \tag{3}$$

Cell State Update: This equation updates the cell state by combining information from the input and forgetting gates [25].

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

Output gate: Responsible for determining which information from the cell state should contribute to generating the output, the output gate is defined by the weight matrices W_o, R_o, and p_o, along with the bias term b_o [25].

$$o_t = \sigma(W_o \cdot x_t + R_o \cdot h_{t-1} + p_o \cdot c_t + b_o) \tag{5}$$

Final output:

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

Bi-LSTMs, as an evolution of standard LSTMs, bolster the performance of models when dealing with sequence classification tasks. In Bi-LSTMs, two LSTM networks are trained on the input sequence rather than just one. The first LSTM processes the original input sequence, while the second processes a reversed copy of the same input sequence. This dual approach is precious when all time steps of the input sequence are available. Fig. 5 visually represents the Bi-LSTM network architecture.

In our model, we incorporated two Bi-LSTM layers. We carefully analyzed the impact of varying the number of LSTM units to identify the optimal configuration. A series of experiments was conducted to balance capturing sufficient temporal dependencies with maintaining computational efficiency. We found that setting the first LSTM units to 512 units and the second to 256 units provided the best performance in terms of both accuracy and training time on our specific dataset.

3.2.3. Multi-head self-attention mechanism

The goal of utilizing the multi-head self-attention mechanism in our Res_Bi-LSTM_MHA model is to improve the model's performance. Models with attention mechanisms often achieve better performance, as they can focus on the most relevant information, reducing noise and improving accuracy. Furthermore, attention weights can provide insights into why the model made specific predictions, making it more interpretable and helping researchers and practitioners understand the decision-making process. Furthermore, multi-head attention is a crucial mechanism that empowers the model to simultaneously focus on information from various representation subspaces and positions [19]. In contrast, using a single attention head can impede such simultaneous processing, as it relies on simple averaging [19]. The multi-head self-attention mechanism consists of two primary components: self-attention and multi-head.

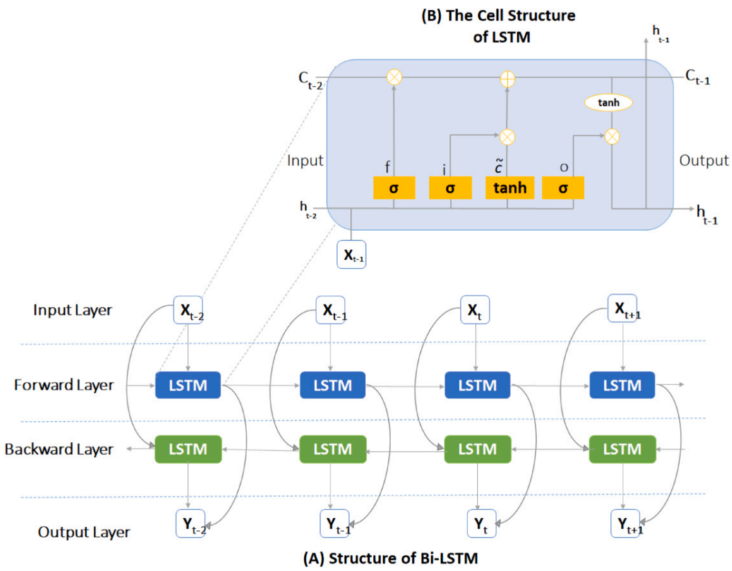

Fig. 5. A Comprehensive exploration of bidirectional long short-term memory (BiLSTM).

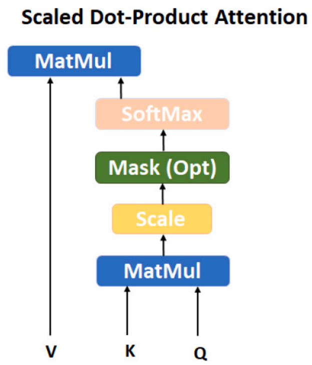

Fig. 6. Scaled dot-product attention.

1. Self-attention:
   The self-attention [19] mechanism enables elements within a sequence to assess their inter-dependencies and determine the extent to which they should prioritize different elements. It does so through a scaled dot-product attention function (Fig. 6). It is one of the attention mechanisms used in the self-attention layers of a transformer model, which operates by mapping a query to a set of key–value pairs to produce an output. This output is a weighted sum of the values, with the weights calculated based on the dot product between the query and the keys. This process can be mathematically formulated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \tag{7}$$

   where V, Q, and K represent the value, query, and key vectors, respectively, and $\sqrt{d_k}$ is the scaling factor, with d_k being the dimension of the key vectors.
   The input includes queries and keys with dimension d_k and values with dimension d_v. We calculate the dot product between the queries and keys, normalize by $\sqrt{d_k}$, and use a softmax function to determine the weights for the values.
2. Multi-head attention:
   Instead of measuring attention only once using d_{model} (dimensional keys, values, and queries), the multi-head mechanism employs scaled dot-product attention multiple times in parallel. Subsequently, 'h' parallel heads are employed to extract information from various channel segments, as shown in Fig. 6. The attention calculation can be described as follows [19]:

$$head_i = \text{attention}(QW_i^O, KW_i^K, VW_i^V) \tag{8}$$

Table 2
Training hyperparameters.

Training parameter	Values	Training parameter	Values
$\beta 1$	0.9	Input Size	(2000,12)
$\beta 2$	0.999	Batch Size	32
Initial Learning Rate	10^{-3}	Training Epochs	60
kernel Size	16	Filters	32
Stride	1	Dropout	0.5

$$\text{Multi-Head}(Q, K, V) = \text{Concat}(head_1, \dots, head_h)W^O \tag{9}$$

where $W_i^O \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^k \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^v \in \mathbb{R}^{d_{\text{model}} \times d_v}$ represent the parameters to be learned, and 'h' stands for the number of heads, the outputs are concatenated and subsequently subjected to linear transformations to achieve the necessary dimensions.

In this study, we utilize four parallel attention layers, meaning the heads are equal to four (h = 4). Each of these heads is equipped with dimensions $d_k = d_v = d_{\text{model}}$ /h = 256. Moreover, we have integrated a concluding fully connected softmax layer, which generates a probability distribution over the nine output classes as indicated in Fig. 4

4. Experimentation and validation

We carried out a series of experiments to evaluate the efficacy of our model. Utilizing multiple metrics and diverse datasets, we assessed its performance in classifying arrhythmias. Furthermore, we compared our results against those achieved by thirty one established methods.

4.1. Experimental setup

The proposed model was developed and trained using Python in conjunction with the Keras deep learning framework, with TensorFlow serving as the backend. Experiments were performed using Google Colab Pro with 83 GB of RAM and GPU acceleration.

4.1.1. Training phase

For model training, the categorical-cross-entropy loss function was utilized, and the Adam optimization method was employed, initially setting the learning rate to 10^{-3}. This learning rate was reduced by a factor of 10 when the loss on the development dataset failed to improve for ten consecutive epochs. To address the significant class imbalance present in the datasets, we adopted the focal loss strategy during training. Focal loss modifies the standard cross-entropy by introducing a modulating factor $(1 - p_t)^\gamma$ that reduces the relative loss for well-classified examples and focuses learning on hard or minority classes. This approach has been shown to be particularly effective for rare arrhythmia types such as PAC and STE. In our experiments, we set $\gamma = 2$, following common practice in ECG classification studies, and found that this helped improve the detection of underrepresented classes without altering the dataset distribution. The hyperparameters $\beta 1$ and $\beta 2$ were set to 0.9 and 0.999, respectively. A mini-batch size of 32 was used during training. The training process spanned 60 epochs, and the model selected for further evaluation was chosen based on its ability to achieve the lowest error on the development dataset. Table 2 provides an overview of the experimental parameters.

In order to thoroughly assess the performance of the proposed method, we conducted experiments using two distinct evaluation strategies: Subject-Wise Evaluation and Record-Wise Evaluation.

In the Record-Wise Evaluation, the proposed model undergoes a thorough assessment across the selected datasets using a comprehensive methodology comprising training, validation, and test sets at the record (ECG segments) level. A 10-fold cross-validation approach is employed, the training datasets are partitioned into ten parts. Nine of these parts are employed for training, while one is reserved for validation. After undergoing several rounds of training and validation, the model demonstrating the most highest performance on the validation dataset is evaluated on unseen test data to gauge its ultimate performance. This approach enables the training of our model on a substantial amount of data, all while maintaining a rigorous evaluation process utilizing independent test and validation sets.

In contrast, in Subject-Wise Evaluation, the data is split at the subject (or patient) level to ensure that no overlap occurs between the training, validation, and test sets. Each set comprises entirely distinct subjects, simulating real-world scenarios where the model must generalize to unseen individuals. By splitting the data at the subject level, we designated randomly 20% of the total dataset as the test set, while the remaining 80% was allocated for training. Of this training data, 20% was further held out as a validation set to monitor the model's performance during training. This approach evaluates the model's ability to generalize to completely new individuals, offering a more realistic assessment of its generalization capabilities.

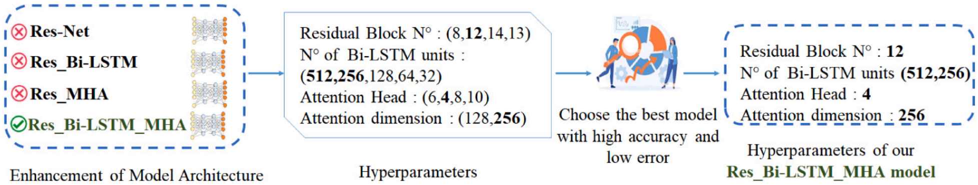

Fig. 7. Optimization and selection of the Res_Bi-LSTM_MHA model architecture and hyperparameters.

4.1.2. Adjustment of model hyperparameters

We employed a random search approach to tune hyperparameters, explicitly tracking performance metrics (accuracy, F1-score) for each configuration to quantify impact, focusing on improving accuracy and reducing error. All experiments used the Record-Wise Evaluation strategy (10-fold cross-validation) for tuning consistency, with final validation on independent subject-wise splits. The Fig. 7 illustrates the different stages involved in the enhancement of the model architecture. Initially, several variants were tested, including Res-Net, Res_Bi-LSTM, Res_MHA, and Res_Bi-LSTM_MHA. After extensive experimentation, the Res_Bi-LSTM_MHA model was found to perform best due to its integration of residual connections, Bi-LSTM units, and multi-head self attention mechanism, which enhanced its ability to handle sequential data and capture long-range dependencies. Key hyperparameters and their performance outcomes are summarized below:

- **Residual Blocks:** Several combinations of residual blocks were tested (e.g., 8, 12, 13, 14), with 12 blocks achieved peak accuracy and F1-score, reducing error by 10% compared to 8 blocks.
- **Bi-LSTM Units:** Through experimentation with different configurations of Bi-LSTM units (512, 256, 128, 64, 32), it was found that using 512 (L1) and 256 (L2) units maximized the performance. Lower units (e.g., 32) dropped accuracy to 72%.
- **Attention Mechanism:** The attention mechanism was also optimized, with the number of heads being set to 4 and the attention dimension to 256. Fewer heads (2) reduced F1-score by 4%.

These adjustments allowed the model to balance complexity and performance, leading to better overall results, such as higher accuracy and reduced error. The proposed Res_Bi-LSTM_MHA model with these hperparameters achieved optimal performance, with an accuracy of 98% and an F1-score of 97% a 10% improvement over baseline methods in F1-score highlighting the critical contribution of the Multi-Head Attention (MHA) mechanism with these adjustments. By combining Bi-LSTM for capturing long-range temporal dependencies, MHA for enhanced feature weighting, and residual connections to preserve gradient flow, the model outperformed alternative approaches by 10%–15% in accuracy (Table 3).

4.2. Ablation experiment

Ablation experiments isolate the contribution of each component in a complex model. In this study, we conduct an ablation study to systematically evaluate the contributions of core components in our Res-Bi-LSTM-MHA model for ECG arrhythmia classification (CPSC-2018 dataset). By selectively disabling individual modules, we isolate their impact on performance, as summarized in Table 3. Results reveal four key insights:

1. **ResNet-only baseline**: The ResNet baseline achieves modest performance (F1-score: 0.76). Establishes the importance of spatial feature extraction for morphological patterns (QRS complexes, ST segments) but shows limited capacity for temporal modeling and global context integration. It is confirmed feature extraction alone is insufficient.
2. **ResNet + Bi-LSTM:** Demonstrates a 11% F1-score improvement over ResNet alone by effectively modeling temporal dependencies and rhythm patterns (e.g., R-R interval variability, arrhythmia progression). Highlights the importance of temporal context.
3. **ResNet + MHA:** (F1 score: 0.81) Confirms MHA's role in capturing long-range dependencies and global context (e.g., ST-T trends across leads) but reveals limitations in modeling sequential dynamics without Bi-LSTM.
4. **Full model (ResNet + Bi-LSTM + MHA):** Achieves a 10% F1-score gain over ResNet-BiLSTM and 16% gain over ResNet-MHA, validating the synergistic combination. This architecture uniquely addresses ECG's multi-scale challenges: ResNet extracts beat-level features, Bi-LSTM models rhythm evolution and MHA integrates distant events and focuses attention.

These findings strongly suggest that the integration of ResNet, Bi-LSTM, and MHA results in a powerful architecture for ECG arrhythmia classification. This combined model surpasses configurations that lack specific components, emphasizing the critical importance of meticulously selecting and integrating diverse architectural elements to attain optimal model performance in the domain of arrhythmia classification.

Table 3
Ablation study results.

ResNet	Performances (%)					
	Bi-LSTM	MHA	Accuracy ↑	Precision↑	Recall ↑	F1-score ↑
✓			0.79	0.75	0.76	0.76
✓	✓		0.88	0.87	0.87	0.87
✓		✓	0.83	0.81	0.81	0.81
✓	✓	✓	**0.98**	**0.97**	**0.98**	**0.97**

Table 4
Layer-wise analysis of Computational complexity, Sequential operations, and Max path length in the Res_Bi-LSTM_MHA Model.

Layer Type	Complexity per Layer	Sequential operations	Max path length
Convolution	$O(n \cdot m^2 \cdot k^2 \cdot d)$	$O(L)$	$O(L)$
Bi-LSTM	$O(2 \cdot n \cdot h^2)$	$O(n)$	$O(n)$
Multi-Head Attention	$O(H \cdot n^2 \cdot d)$	$O(1)$	$O(1)$

4.3. Evaluation measures

The metrics used in our evaluation included overall accuracy (Acc), precision (P), recall rate (R), and the F1 score. Detailed descriptions of these metrics are provided below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

$$\text{Macro avg F1-score} = \frac{1}{N} \sum_{i=1}^{N} \text{F1-score}_i \tag{14}$$

In this context, TP represents the count of correctly classified samples within this class, FN stands for the count of samples from this class that were misclassified as belonging to other classes, and FP represents the count of samples misclassified as belonging to this class when they belong to other classes. The F1 score for all nine categories was averaged to assess the model's overall performance.

4.4. Computational complexity

We present a comprehensive comparison of the computational complexity (CC) per layer for the proposed Res_Bi-LSTM_MHA model, along with its individual components in Table 4. The Bi-LSTM layer, which is inherently sequential, requires $O(n)$ sequential operations and has a path length of $O(n)$. In contrast, the multi-head attention layer capitalizes on its parallelization capabilities, requiring only $O(1)$ sequential operations and achieving a path length of $O(1)$. From a time and space complexity standpoint, the multi-head attention layer demonstrates a lower computational burden than both the convolutional and Bi-LSTM layers, following the complexity hierarchy:

$$CC_{\text{multi-head attention}} < CC_{\text{convolution}} < CC_{\text{Bi-LSTM}} \tag{15}$$

While the convolutional layers increase the overall complexity of the Res_Bi-LSTM_MHA model by focusing on local feature extraction, the inclusion of multi-head attention greatly improves computational efficiency by reducing both time and space complexity compared to CNN or CNN-LSTM architectures. This integration allows the model to effectively capture both local and global dependencies, striking a balance between computational complexity and performance.

4.5. Obtained results using MIT-BIH dataset

In this section, we present the performance of our model on the MIT-BIH Arrhythmia Database, a well-established benchmark for ECG arrhythmia classification. Given its controlled environment and standardized annotations, this dataset serves as a foundation for assessing the baseline capabilities of our model. The results outlined below reflect the model's effectiveness in detecting various arrhythmia types, both in record-wise and subject-wise evaluations.

Table 5
Res_Bi-LSTM_MHA model performances on the testing dataset using the MIT-BIH dataset.

Performances	MIT-BIH Dataset								
	Record Wise Evaluation					Subject Wise Evaluation			
	NSR	LBBB	RBBB	PAC	PVC	NSR	LBBB	RBBB	PVC
Precision ↑	100%	99%	100%	96%	99%	97%	85%	84%	100%
Recall ↑	100%	100%	100%	90%	98%	94%	99%	62%	100%
F1-score ↑	100%	99%	100%	93%	98%	95%	91%	72%	100%
Accuracy ↑	99.42%					96.44%			
Avg F1 ↑	98.01%					90%			

In this section, we present the performance of our model on the MIT-BIH Arrhythmia Database, a well-established benchmark for ECG arrhythmia classification. Given its controlled environment and standardized annotations, this dataset serves as a foundation for assessing the baseline capabilities of our model. The results outlined below reflect the model's effectiveness in detecting various arrhythmia types, both in record-wise and subject-wise evaluations. As shown in Table 5, our model demonstrated strong performance in detecting different arrhythmia types. In the record-wise evaluation, the model achieved an impressive accuracy of 99.42% and an F1-score of 98.01%. It exhibited nearly perfect precision and recall, particularly for NSR, RBBB, and LBBB classes, all reaching 100%. While the PAC class shows a slightly lower F1-score of 93%, this can be attributed to the limited number of instances within this particular class. Notably, its precision of 96% and recall of 90% evaluation indicate strong predictive capability for this minority class, though the recall gap suggests potential missed diagnoses.

In the subject-wise evaluation, the model performance showed variability in performance, especially for RBBB and PAC. The precision–recall tradeoff is particularly evident for RBBB (84% precision vs 62% recall) and PAC (84% precision vs 72% recall), highlighting challenges in generalizing these arrhythmia detections to new patients. However, the model maintained near-perfect precision and recall for NSR, LBBB, and PVC, confirming its reliability for these conditions. With an average F1-score of 90% and overall accuracy of 96.44%, the model remains highly effective, though individual-level evaluation presents more complexity and room for improvement. It is also important to highlight the performance on minority classes. In both evaluation settings, the PAC class, which is relatively rare in the MIT-BIH dataset, achieved lower recall (90% in record-wise and 72% in subject-wise evaluation), indicating that some abnormal beats of this type were missed. Similarly, the RBBB class showed reduced recall (62% subject-wise), reflecting challenges in reliably identifying these less frequent arrhythmias when applied to unseen subjects. These findings emphasize that while the overall averages remain high, the detection of underrepresented classes is more difficult. The use of focal loss in our training procedure partially alleviated this issue by emphasizing difficult-to-classify examples, leading to competitive precision and improved balance across classes. Nevertheless, the performance gap for rare classes highlights a key limitation and suggests that future work should explore complementary imbalance-handling strategies such as targeted oversampling, synthetic beat generation, or cost-sensitive learning.

It is important to note that record-wise evaluation can lead to inflated performance metrics because ECG segments from the same subject may appear in both training and testing folds, allowing the model to exploit subject-specific patterns. Subject-wise evaluation, which more accurately reflects clinical deployment, shows a noticeable performance drop (average F1 from 98.01% to 90%), particularly in classes such as RBBB and PAC. This indicates challenges in generalizing across individuals with distinct physiological characteristics.

4.6. Obtained results using INCART dataset

The St Petersburg INCART 12-lead arrhythmia database offering a more diverse patient population and includes 12-lead ECG recordings, creating a greater challenge for arrhythmia detection. Evaluating our model on this dataset enables us to assess its generalization capabilities across more complex and varied data. The following results illustrate the performance of the model in both record-wise and subject-wise evaluations.

Table 6 presents a thorough analysis of the classification performance of the Res_Bi-LSTM_MHA model on the test data of the INCART dataset, evaluated in both the record-wise and subject-wise scenarios. The performance of the model in the record-wise evaluation demonstrates exceptional results, with a precision rate of 99% to detect NSR, PVC, and PAC, and an exceptional precision of 100% for RBBB. This indicates the model's excellent capacity to correctly identify arrhythmias, minimizing the occurrence of false positives rate. Recall scores are equally impressive, with perfect 100% for NSR and PVC, and almost 100% for other arrhythmias such as RBBB and PAC. These high recall rates suggest the model rarely misses arrhythmia cases, which is essential in clinical settings where the accurate detection of such conditions is critical. The F1-scores, exceed 95% for all categories, with a flawless 100% for RBBB, underscoring the robustness of the model. Furthermore, the model's overall accuracy of 99.24% further reinforces its ability to generalize effectively across the testing data.

In contrast, the subject-wise evaluation reveals slight variations in precision and recall, particularly for PVC, where the precision decreases to 86%, although the recall remains high at 96%. The precision–recall disparity for PVC suggests our model prioritizes sensitivity over specificity for this condition a clinically desirable behavior given PVC's diagnostic significance. PAC exhibits a lower recall of 87%, which affects its F1-score, reducing it to 93%. These fluctuations likely result from inter-patient variability, which

Table 6
Res_Bi-LSTM_MHA model performances on the testing dataset using the INCART dataset.

Performances	INCART Dataset						
	Record Wise Evaluation				Subject Wise Evaluation		
	NSR	PVC	PAC	RBBB	NSR	PVC	PAC
Precision ↑	99%	99%	99%	100%	98%	86%	99%
Recall ↑	100%	100%	93%	99%	94%	96%	87%
F1-score ↑	99%	99%	96%	100%	96%	91%	93%
Accuracy ↑	99.24%				94.08%		
Avg F1 ↑	98.71%				93.33%		

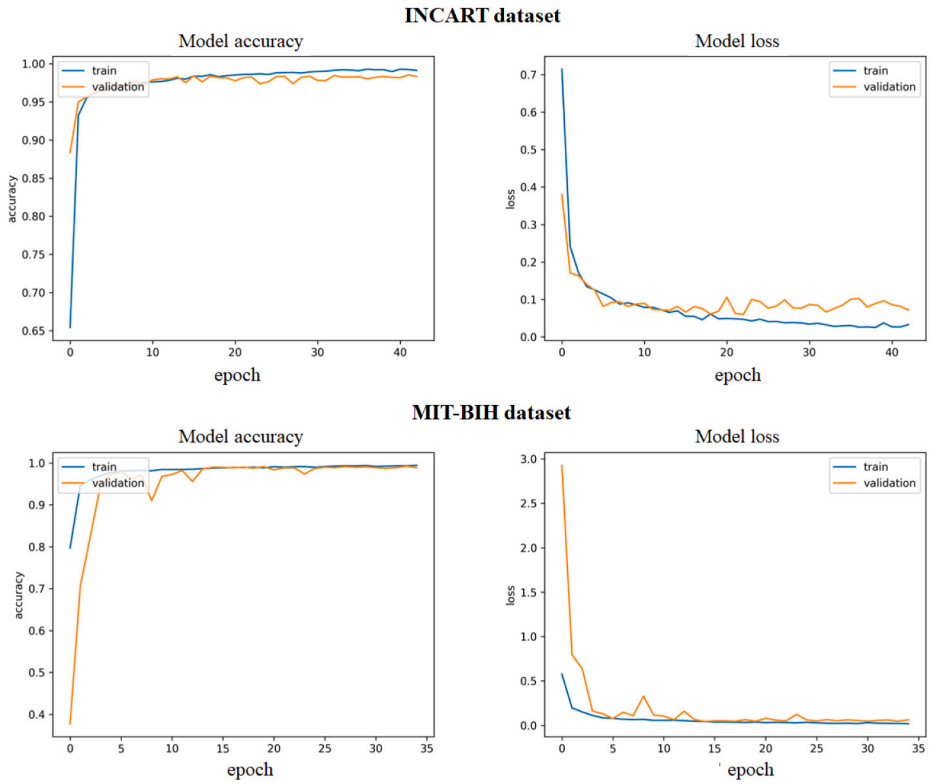

Fig. 8. Accuracy and loss curves for INCART and MIT-BIH datasets.

makes subject-wise evaluations more complex and challenging. Despite this variability, the model achieves an average F1-score of 93.33% and an accuracy of 94.08%, indicating a high degree of generalizability even in the face of patient differences.

We present the accuracy and loss curves derived from the model's training and validation phases on both the INCART and MIT-BIH datasets in Fig. 8. The absence of overfitting is evident, as the validation curve closely mirrors the trajectory of the training curve. The curves exhibit steady improvement throughout the iterations, suggesting consistent enhancement over time. This alignment indicates the model's successful generalization of new data, leading to reliable performance on the training and validation sets.

4.7. Obtained results using CPSC 2018 dataset

To further demonstrate the broad applicability of our proposed Res_Bi-LSTM_MHA model, we conducted additional classification experiments using the CPSC 2018 Dataset. The CPSC 2018 dataset presents a valuable yet challenging benchmark for evaluating the performance of ECG classification models. Its complexity stems from several factors, such as the multiclass nature of arrhythmia classification, significant class imbalance, and the inherent variability in ECG signals. Despite these difficulties, the dataset is widely adopted in the literature due to its clinical relevance and the broad spectrum of arrhythmia types it represents. However, enhancing the accuracy of classification continues to pose a significant challenge. This makes it an ideal testbed for assessing the accuracy and robustness of our model, particularly in addressing challenges like rare arrhythmia detection and noisy signals. By leveraging the CPSC 2018 dataset, we ensure that our model undergoes a rigorous evaluation under realistic, clinically applicable conditions,

Table 7
Res_Bi-LSTM_MHA model performances on the testing dataset using the CPSC 2018 Dataset.

Categories	CPSC 2018 Dataset					
	Record Wise Evaluation			Subject Wise Evaluation		
	Precision ↑	Recall ↑	F1-score ↑	Precision ↑	Recall ↑	F1-score ↑
NSR	98%	100%	99%	90%	99%	94%
AF	96%	99%	97%	94%	100%	97%
I-AVB	96%	96%	96%	100%	92%	96%
LBBB	94%	100%	97%	97%	100%	98%
RBBB	99%	97%	98%	99%	97%	98%
PAC	96%	97%	96%	95%	85%	90%
PVC	99%	97%	98%	100%	94%	97%
STD	98%	98%	98%	95%	97%	96%
STE	100%	97%	98%	98%	94%	96%
Accuracy ↑		98%			95.81%	
Avg F1 ↑		97%			95.87%	

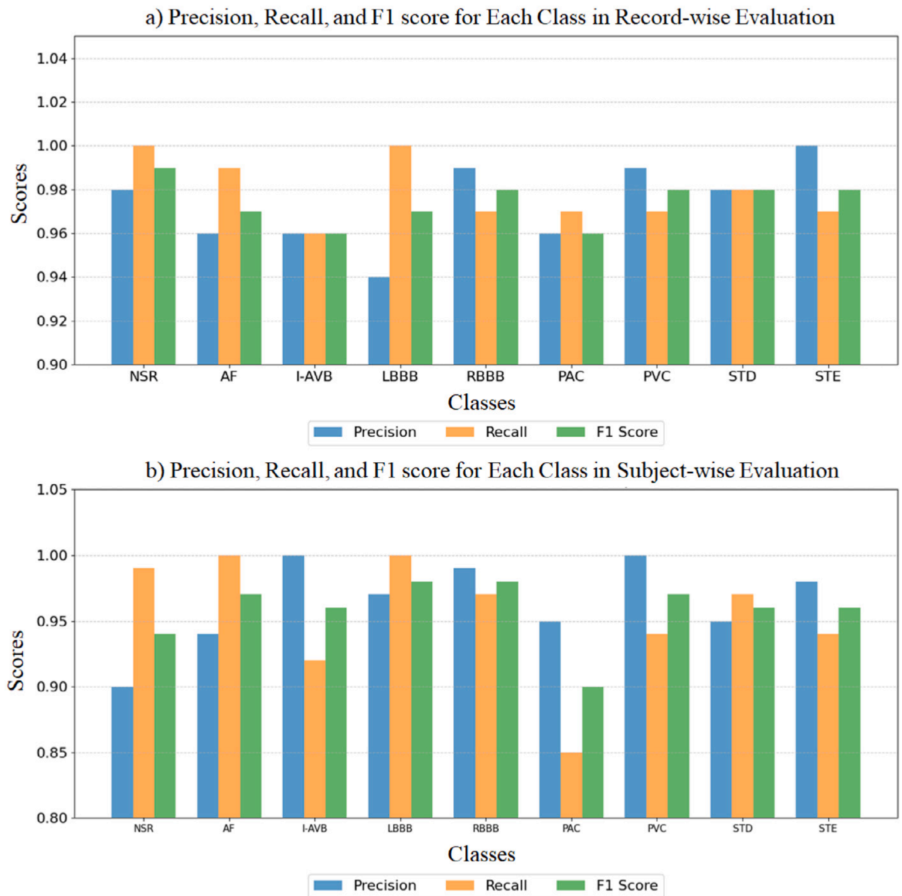

Fig. 9. Per-Class Metrics in Record- and Subject-Wise Evaluation on CPSC 2018.

paving the way for advancements in ECG-based arrhythmia detection. The results of this evaluation are presented in Table 7 and Fig. 9.

Table 7 provides a comprehensive overview of the model's performance on the testing dataset, highlighting its robustness and generalization capabilities across different evaluation methods. The model achieves an overall accuracy of 98% in the record-wise evaluation, which slightly decreases to 95.81% in the subject-wise evaluation. This decline can be attributed to the inherent complexity of subject-wise evaluation, where the model must generalize across unseen subjects, underscoring its ability to maintain high performance even under more challenging conditions.

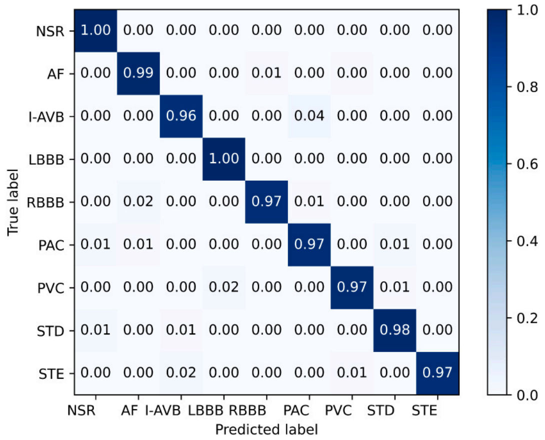

Fig. 10. Res_Bi-LSTM_MHA confusion matrix for nine-class arrhythmia classification on testing data.

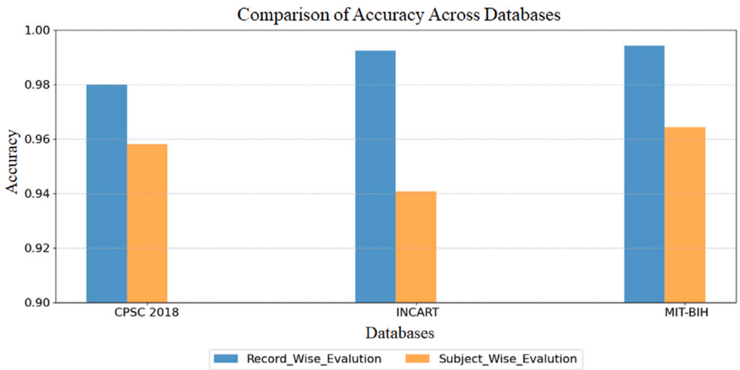

Fig. 11. Comparative analysis of accuracy scores across three datasets (The CPSC 2018, The INCART, and The MIT-BIH).

Across all categories, the model consistently exhibits strong performance, with F1-scores ranging from 90% to 98% in the subject-wise evaluation and from 96% to 99% in the record-wise evaluation. Notably, categories such as NSR, RBBB, PVC, STD, and STE achieve high F1-scores, ranging from 98% to 99% in the record-wise evaluation. In contrast, other categories such as AF, IAVB, and LBBB, while still performing well, show slightly lower F1-scores, approximately ranging from 96% to 97% in the record-wise evaluation. This consistency across diverse classifications emphasizes the model's high precision and recall. Moreover, the model demonstrates exceptional performance in detecting NSR, AF, IAVB, LBBB, RBBB, PVC, STD, and STE, with F1-scores ranging from 94% to 98% in the subject-wise evaluation. While STE maintains high precision (98%), its 94% recall warrants attention, as missed ST-elevation cases could delay critical STEMI interventions. Furthermore, a slight drop in performance is observed in certain arrhythmia types, such as PAC, where the F1-score decreases from 96% in the record-wise evaluation to 90% in the subject-wise evaluation. This drop can be attributed to the class imbalance in the dataset and the relatively small number of PAC samples, which makes the model's task more challenging.

These results underscore the model's potential for reliable arrhythmia detection in clinical settings. Overall, the Res_Bi-LSTM_MHA model demonstrates a high level of accuracy and consistency in arrhythmia classification, offering promising results for both record-wise and subject-wise evaluations. These findings validate the model's applicability in real-world scenarios, where accurate and reliable ECG classification is crucial.

The confusion matrix stands as a valuable tool for evaluating the efficacy of an ECG classification model. In Fig. 10, we present the confusion matrix illustrating the test results of our Res_Bi-LSTM_MHA model using record-wise evaluation. This matrix offers a comprehensive view of the model's performance across various categories of cardiac arrhythmias. Remarkably, our model exhibits exceptional accuracy in classifying all arrhythmia types, underlining its effectiveness in accurately distinguishing between different arrhythmias.

The overall results across the CPSC, INCART and MIT-BIH datasets demonstrates consistently high performance in both record-wise and subject-wise evaluations (As shown in Fig. 11). Record-wise evaluation consistently yields higher accuracy scores across all datasets. However, a slight drop in accuracy is observed in the subject-wise evaluation, reflecting some challenges in maintaining generalization across different subjects. Nonetheless, the overall performance remains strong across both evaluation strategies, indicating the model's effectiveness and potential applicability across various datasets for arrhythmia classification.

Table 8
Comparative analysis of arrhythmia classification using various models on the CPSC-2018 dataset.

Method	Split	Accuracy↑	Precision↑	Recall↑	F1-score↑
CNN_Transformer_LC [10]	–	–	–	–	78.6%
AFibNet [15]	7409 train 823 validation 3610 test	96.36%	92.47%	–	92.94%
ConvLSTM [3]	–	84%	–	–	–
SPN-V2 [2]	10-fold CV	83.6%	79.4%	78.5%	78.7%
CNN-RNN [12]	10-fold CV	77.4%	83.9%	73.4%	77.3%
ResRNN [1]	–	91%	91%	91%	91%
DAMS-Net [20]	80% train 10% validation 10% test	–	84.7	83.7	83.9
MTGBI-LSTM [16]	5-fold CV	96.48%	–	–	–
ParNet-adv [17]	5-fold CV	96.74%	95.18%	98.75%	96.93%
CNN-RNN-attention [27]	70% train 10% validation 20% test	–	–	–	90.82%
CNN-DVIT [28]	10-fold CV	–	81.9%	84.9%	82.9%
CNN+Transformer [11]	10-fold CV	82.7%	–	–	84.7%
CNN-Transformer+ external attention mechanism [22]	5-fold CV	86.3%	–	–	85.4%
Res_Bi-LSTM_MHA	**10-fold CV**	**98%**	**97.33%**	**98%**	**97.45%**

4.8. Performance comparison with existing methods

In this subsection, we present a comparative analysis of our results with existing methods in the literature. Table 8 reveals that the arrhythmia classification results achieved with the proposed Res_Bi-LSTM_MHA method consistently outperform those of various comparison models utilizing the CPSC-2018 [8] dataset across a majority of the assessed factors. Our model significantly enhances accuracy, recall, precision, and F1 scores. It is worth highlighting that in the testing set, the experimental results of the Res_Bi-LSTM_MHA model surpass those of models based on Convolutional Neural Network Modules (ResRNN [1], SPN-V2 [2],Con-vLSTM [3], AFibNet [15], MTGBI-LSTM [16], ParNet-adv [17], CNN-RNN [12]) and Convolutional Neural Network Modules with Attention Mechanisms (CNN_Trans- former_LC [10], CNN_Transformer [11], DAMS-Net [20], CNN-Transformer+external attention mechanism [22], CNN-RNN-attention [27], CNN-DVIT [28]).

Furthermore, we conducted a comprehensive comparative analysis of nine types of arrhythmias (Normal and Abnormal), contrasting our Res_Bi-LSTM_MHA with recently introduced 12-lead ECG classification methods. This analysis encompasses both fundamental neural networks and attention mechanisms, and it was performed on the CPSC 2018 dataset [8]. As illustrated in Table 9, previous studies predominantly utilized conventional neural network architectures such as CNN, Attention, and Transformer for ECG classification, achieving commendable performance levels. To further bolster efficacy and explore novel avenues, this paper introduces an innovative ResNet+Bi-LSTM+Multi-Head Attention framework. This framework seamlessly integrates convolutional neural networks, recognized for their proficiency in extracting local features, with attention mechanisms renowned for processing sequence data and capturing global dependencies. As demonstrated in Table 9, our Res_Bi-LSTM_MHA boasts an impressive average F1 score of 97%, outperforming the other methods by a significant margin. Our model secured the highest F1 score in eight out of the nine ECG categories. Specifically, our approach surpassed the CNN Transformer LC [10], DAMS-Net [20], CNN-RNN-attention [27], CNN-DVIT citedong2023arrhythmia, CNN-Transformer [11], and CNN-Transformer+external attention mechanism [22] methods in all nine categories, with an average F1 score advantage of 0.184, 0.06, 0.061, 0.141, 0.11, and 0.11 respectively. Furthermore, our F1 scores in all categories outperformed the other methods, except for LBBB, where ResRNN [1] achieved a slightly higher score. Notably, in direct comparison to the most recent competitor [28], which also incorporated the Multi-Head Attention, our Res_Bi-LSTM_MHA excelled in each category, with an average F1 score improvement of 0.141.

Fig. 12 illustrates the performance of the Res_Bi-LSTM_MHA model in comparison to existing studies that employ convolutional neural networks with attention mechanisms for 12-Lead ECG classification using the CPSC 2018 dataset. The results demonstrate that our model significantly outperforms the latest competitors in the field.

To showcase the broad applicability of our proposed Res_Bi-LSTM_MHA model, we performed supplementary classification experiments using the INCART [9] and the MIT-BIH arrhythmia [5] datasets, as outlined in Table 10 the experimental configurations were consistent with those previously described, employing Res_Bi-LSTM_MHA for ECG signal classification. The outcomes demonstrate that our Res_Bi-LSTM_MHA model attains a state-of-the-art F1-score of 98.01% and accuracy of 99.42% on the MIT-BIH arrhythmia database [5], surpassing the performance of numerous published models. In contrast to ResNet, BI-LSTM, and Attention methodologies, our model achieved significantly superior F1-score and accuracy. This highlights the benefits of concurrently extracting features through ResNet, BI-LSTM, and a Multi-head self-attention mechanism and provides additional confirmation of the generalization prowess of the Res_Bi-LSTM_MHA model.

In the comparative analysis of arrhythmia classification methods on the INCART database, various models have been scrutinized for their performance, each offering unique insights into the challenges of ECG Arrhythmia Classification as summarized in Table 11.

Table 9
Classification performance of nine types of arrhythmias (normal and abnormal) through comparative analysis across various studies in the literature.

Methods	Type									Average F1↑
	NSR	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE	
CNN_Transformer_LC [10]	81.7%	85.8%	87.8%	80%	87%	61.8%	83%	71.1%	68.6%	78.6%
ResRNN [1]	90%	94%	89%	**98%**	96%	80%	89%	81%	96%	91%
DAMS-Net [20]	81.9%	91.5%	88.1%	87.8%	93.6%	75.5%	87.6%	81.9%	68.4%	83.9%
CNN-RNN-attention [27]	92.13%	–	–	90.49%	90.67%	90.30%	90.51%	–	–	90.82%
CNN-DVIT [28]	83.1%	92.4%	88.7%	90.5%	93.5%	70.4%	84.2%	82.3%	61%	82.9%
CNN+Transformer [11]	89%	94%	88%	87%	94%	80%	90%	87%	65%	86%
CNN-Transformer + external attention mechanism [22]	81.3%	90.8%	86.3%	91.8%	92.4%	76.8%	83.3%	82.7%	82.8%	85.4%
Proposed Method	**99%**	**97%**	**96%**	97%	**98%**	**96%**	**98%**	**98%**	**98%**	**97%**

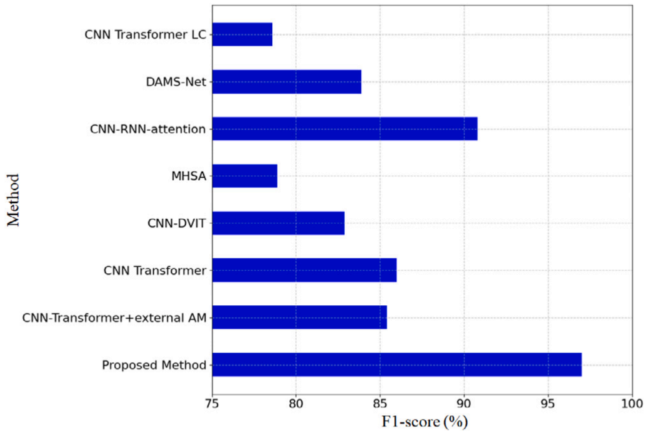

Fig. 12. Comparative analysis of arrhythmia classification in studies utilizing attention mechanisms.

Supriya et al. [29] demonstrated a commendable accuracy of 94.1%, underscoring the robustness of convolutional neural networks in this domain. Conversely, their DNN counterpart exhibited a slightly lower accuracy of 87.7%. Chen et al. [30] proposed a CMM model, achieving an impressive F1-score of 85.12%, emphasizing the importance of a balanced precision–recall trade-off. Vasconcellos et al. [31] and Park et al. [32] introduced notable contributions with accuracies of 86% and 92.47%, respectively. Gupta et al. [33] utilized data programming and augmentation to attain an accuracy of 89.4%, shedding light on the impact of augmentation strategies. However, the pinnacle of this comparative analysis is the proposed Res_Bi-LSTM_MHA model, showcasing a remarkable accuracy of 96% and an F1-score of 93.6%. This underscores the efficacy of our Res_Bi-LSTM_MHA architecture in significantly advancing the state-of-the-art in 12-lead ECG arrhythmia classification.

4.9. Threats to validity

While our proposed model demonstrates strong performance across all datasets, several important threats to validity must be acknowledged. These relate primarily to evaluation protocols, generalization to new subjects, and dataset characteristics.

1. **External Validity (Generalizability Concerns)**
   A key external validity threat arises from the discrepancy between record-wise and subject-wise evaluations. As presented in Table 5, record-wise evaluation consistently yields higher performance metrics. This is partly due to **data leakage**, as ECG segments from the same subject may appear in both training and testing sets. In such cases, the model can exploit subject-specific patterns (e.g., baseline morphology, heart rate dynamics) rather than learning generalizable arrhythmia representations. This leads to an optimistic estimation of performance and may not reflect real-world clinical deployment.
   Subject-wise evaluation, by contrast, provides a more realistic scenario where models are tested on completely unseen patients. Here, we observe a clear performance gap—for instance, the average F1-score drops from 98.01% (record-wise) to 90% (subject-wise) on the MIT-BIH dataset. Similar trends are observed for INCART and CPSC 2018. This drop highlights the challenges of inter-patient variability, including differences in ECG morphology, signal amplitude, electrode placement, and noise characteristics. These factors make generalization to new individuals more difficult, which is a critical consideration for clinical applicability.

Table 10
Comparative analysis of arrhythmia classification using various models on the MIT-BIH dataset.

Method	Split	Classes	Accuracy↑	F1-score↑
RBF [14]	40% train 25% validation 35% test	2	98.79%	–
ConvLSTM [3]	–	2	84%	–
ResNet-18 [4]	5-fold CV	4	97.21%	95.96%
DeepArr [6]	10-fold CV	5	**99.46%**	97.63%
CNN-LSTM Attention [7]	5-fold CV	5	98.95%	–
MTGBi-LSTM [16]	5-fold CV	5	99.2%	94%
CNN-RNN-Attention [27]	70% train 10% validation 20% test	5	84.65%	91.10%
CMM [30]	–	3	–	82.93%
LDCNN [18]	*	5	99.38%	**99.6%**
DP+DA [33]	–	2	92.2%	95.4%
CNN+Transformer [11]	10-fold CV	–	99.28%	–
CAT-Net [21]	80% train 20% test	5	99.14%	94.69%
Res_Bi-LSTM_MHA	**10-fold CV**	5	**99.42%**	**98.01%**

Table 11
Comparative analysis of arrhythmia classification using various models on the INCART dataset.

Method	Accuracy↑	F1-score↑
CNN [29]	94.1%	–
DNN [29]	87.7%	–
CMM [30]	–	85.12%
SCNN [31]	86%	–
LSTM-FCN with self-attention [32]	92.47%	–
DP+DA [33]	89.4%	–
PVCNet [34]	98.50%	93.19%
Res_Bi-LSTM_MHA	**99.24%**	**98.71%**

2. **Failure Cases and Class-wise Performance.** To better understand this gap, we examined class-wise results under subject-wise evaluation. The most affected categories were those with fewer training samples or higher waveform variability. For example, on MIT-BIH, the RBBB class shows a recall drop to 62%, and PAC also experiences a notable decline. These errors typically arise from subtle inter-patient morphological differences or noisy recordings, such as baseline drift and motion artifacts. Such failure patterns indicate that the model struggles to generalize rare or morphologically variable arrhythmias across different subjects, even when overall accuracy remains high.
3. **Internal Validity Considerations** Internal validity is influenced by the data partitioning strategy. While subject-wise splitting reduces data leakage, the limited number of unique subjects in some datasets restricts statistical power and increases variability in results. Additionally, hyperparameter tuning was primarily conducted using record-wise splits to ensure sufficient training data, which may introduce slight biases when transferring to subject-wise evaluation.

To address these threats, we propose several directions for future work:

- **Advanced data Augmentation:** Target physiological variability using synthetic noise injection and time warping to improve robustness to inter-patient variability.
- **Transfer Learning:** Pre-train on larger and more diverse datasets (e.g., PTB-XL, CPSC 2018) and fine-tune on smaller target datasets to enhance generalization.
- **Metadata-Aware Normalization:** Incorporate patient metadata (e.g., age, sex, BMI) into normalization layers to adapt the model to demographic differences.
- **Targeted Evaluation:** Include class-wise and failure mode analyses systematically, focusing on rare arrhythmias and challenging cases to identify weaknesses more effectively.

5. Conclusion

The critical nature of arrhythmia detection in modern healthcare demands robust computational tools capable of surpassing the limitations of traditional diagnostic methods. This research addressed these challenges through the development of the Res_Bi-LSTM_MHA architecture. By integrating the feature extraction capabilities of ResNet with the temporal depth of Bi-LSTM and the selective focus of multi-head self-attention, the model effectively captures the complex nuances inherent in electrocardiogram (ECG) data.

The experimental validation across three benchmark datasets, CPSC 2018, INCART, and MIT-BIH, confirms the high performance and versatility of this hybrid approach. The model consistently outperformed alternative methods, achieving an accuracy of 98% and an average F1 score of 97% on the CPSC 2018 dataset. Its strong performance across the diverse 12-lead signals of INCART and the classic ambulatory recordings of MIT-BIH further demonstrates its reliability in handling varied clinical data sources and signal complexities. These results suggest that the synergistic combination of residual learning and attention mechanisms is highly effective for processing high-dimensional physiological signals.

Moving forward, the focus will shift toward enhancing the model's adaptability and clinical utility. First, to improve generalization across broader patient populations, a transfer learning framework will be developed using larger repositories such as the Chapman-Shaoxing and PTB-XL datasets. This will involve pre-training on expansive datasets to better account for rare arrhythmias and inter-patient variability. Second, to facilitate real-time clinical adoption, the implementation of a high-speed Grad-CAM visualization framework—targeting an update rate of at least 250 Hz—will be prioritized. This will ensure that the model's predictions are not only accurate but also interpretable for clinicians in rapid diagnostic settings.

CRediT authorship contribution statement

Saoueb Kerdoudi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Validation, Writing – original draft. **Larbi Guezouli:** Supervision, Methodology, Writing – review & editing. **Tahar Dilekh:** Supervision, Methodology, Writing – review & editing.

Declaration of competing interest

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.

Data availability

The MIT-BIH and INCART datasets are publicly available on PhysioNet. The CPSC2018 dataset is also available via PhysioNet as part of the PhysioNet/Computing in Cardiology Challenge 2020 training data.

References

[1] Dhyani S, Kumar A, Choudhury S. Arrhythmia disease classification utilizing resrnn. Biomed Signal Process Control 2023;79:104160.
[2] Huang Y, Yen GG, Tseng VS. Snippet policy network v2: Knee-guided neuroevolution for multi-lead ecg early classification. IEEE Trans Neural Netw. Learn Syst 2022.
[3] Hammad M, Abd El-Latif AA, Hussain A, Abd El-Samie FE, Gupta BB, Ugail H, Sedik A. Deep learning models for arrhythmia detection in IoT healthcare applications. Comput Electr Eng 2022;100:108011.
[4] Mathunjwa BM, Lin Y-T, Lin C-H, Abbod MF, Sadrawi M, Shieh J-S. ECG recurrence plot-based arrhythmia classification using two-dimensional deep residual CNN features. Sensors 2022;22(4):1660.
[5] Moody GB, Mark RG. The impact of the MIT-bih arrhythmia database. IEEE Eng Med Biol Mag 2001;20(3):45–50.
[6] Midani W, Ouarda W, Ayed MB. DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals. Biomed Signal Process Control 2023;85:104954.
[7] Huang Y, Li H, Yu X. A novel time representation input based on deep learning for ecg classification. Biomed Signal Process Control 2023;83:104628.
[8] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, Liu Y, Ma C, Wei S, He Z, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J Med Imaging Health Informat. 2018;8(7):1368–73.
[9] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St petersburg INCART 12-lead arrhythmia database. PhysioBank PhysioToolkit PhysioNet 2008. [Online; accessed YYYY-MM-DD].
[10] Che C, Zhang P, Zhu M, Qu Y, Jin B. Constrained transformer network for ECG signal processing and arrhythmia classification. BMC Med Inform Decis Mak 2021;21(1):1–13.
[11] Ji C, Wang L, Qin J, Liu L, Han Y, Wang Z. Msgformer: A multi-scale grid transformer network for 12-lead ecg arrhythmia detection. Biomed Signal Process Control 2024;87:105499.
[12] Yang J, Li J, Lan K, Wei A, Wang H, Huang S, Fong S. Multi-label attribute selection of arrhythmia for electrocardiogram signals with fusion learning. Bioengineering 2022;9(7):268.
[13] Habibi Z, Karimizadeh K, Nosratpour A, Alipourfard I. Enhanced QRS detection and ECG compression using adaptive thresholding: a real-time approach for improved monitoring and diagnosis. Comput Electr Eng 2024;119:109528.
[14] Mateo J, Torres AM, Aparicio A, Santos JL. An efficient method for ECG beat classification and correction of ectopic beats. Comput Electr Eng 2016;53:219–29.
[15] Tutuko B, Nurmaini S, Tondas AE, Rachmatullah MN, Darmawahyuni A, Esafri R, Firdaus F, Sapitri AI. AFibNet: an implementation of atrial fibrillation detection with convolutional neural network. BMC Med Inform Decis Mak 2021;21:1–17.
[16] Munawar S, Angappan G, Konda S. Arrhythmia classification based on bi-directional long short-term memory and multi-task group method. Int J E-Collaboration (IJeC) 2023;19(1):1–18.
[17] Zhang H, Liu C, Tang F, Li M, Zhang D, Xia L, Crozier S, Gan H, Zhao N, Xu W, et al. Atrial fibrillation classification based on the 2D representation of minimal subset ECG and a non-deep neural network. Front Physiol 2023;14:182.
[18] Bayani A, Kargar M. LDCNN: A new arrhythmia detection technique with ECG signals using a linear deep convolutional neural network. Physiol Rep 2024;12(17):e16182.
[19] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[20] Zhou R, Yao J, Hong Q, Zheng Y, Zheng L. DAMS-net: Dual attention and multi-scale information fusion network for 12-lead ECG classification. Methods 2023;220:134–41.

[21] Islam MR, Qaraqe M, Qaraqe K, Serpedin E. Cat-net: Convolution, attention, and transformer based network for single-lead ecg arrhythmia classification. Biomed Signal Process Control 2024;93:106211.
[22] Li H, Han J, Zhang H, Zhang X, Si Y, Zhang Y, Liu Y, Yang H. Clinical knowledge-based ECG abnormalities detection using dual-view CNN-transformer and external attention mechanism. Comput Biol Med 2024;178:108751.
[23] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.
[24] Reddy SR, Varma GS, Davuluri RL. Resnet-based modified red deer optimization with DLCNN classifier for plant disease identification and classification. Comput Electr Eng 2023;105:108492.
[25] Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev 2020;53:5929–55.
[26] Rahul J, Sharma LD. Automatic cardiac arrhythmia classification based on hybrid 1-D CNN and bi-LSTM model. Biocybern Biomed Eng 2022;42(1):312–24.
[27] Sun J. Automatic cardiac arrhythmias classification using CNN and attention-based RNN network. Heal Technol Lett 2023.
[28] Dong Y, Zhang M, Qiu L, Wang L, Yu Y. An arrhythmia classification model based on vision transformer with deformable attention. Micromachines 2023;14(6):1155.
[29] Supriya M, Shankar SP, Reddy SSM, Geethanjali R, Sathvika S, Shetty S. Classification of ventricular arrhythmia using machine learning and deep learning techniques. In: 2022 4th international conference on circuits, control, communication and computing (i4C). IEEE; 2022, p. 36–41.
[30] Chen L, Lian C, Zeng Z, Xu B, Su Y. Cross-modal multiscale multi-instance learning for long-term ecg classification. Inform Sci 2023;119230.
[31] Vasconcellos ME, Ferreira BG, Leandro JS, Neto BF, Cordeiro FR, Cestari IA, Gutierrez MA, Sobrinho Á, Cordeiro TD. Siamese convolutional neural network for heartbeat classification using limited 12-lead ECG datasets. Ieee Access 2023;11:5365–76.
[32] Park J, Lee K, Park N, You SC, Ko J. Self-attention LSTM-FCN model for arrhythmia classification and uncertainty assessment. Artif Intell Med 2023;142:102570.
[33] Gupta P, Nandakumar S, Gupta M, Panda G. Data programming enabled weak supervised labeling for ECG time series. Biomed Signal Process Control 2024;87:105540.
[34] Wang Z, Wang K, Chen X, Zheng Y, Wu X. A deep learning approach for inter-patient classification of premature ventricular contraction from electrocardiogram. Biomed Signal Process Control 2024;94:106265.

Saoueb Kerdoudi is a Ph.D. student at the University of Batna 2, Algeria. Her research interests include signal processing, AI, machine learning, health monitoring, and smart homes. She has published several papers in international conferences and is active in real-time systems and cloud computing research.

Larbi Guezouli graduated in computer science from the University of Paris 7, France, in 2007. He is currently a teacher-researcher at the Higher National School of Renewable Energy, Batna, Algeria. He has published numerous articles in his fields of interest, which include information retrieval, data mining, machine learning, artificial intelligence, big data, and IoT.

Tahar Dilekh received his PhD in Computer Science from University of Batna 2 in 2019. He is currently a lecturer at the same University, where his research focuses on machine learning and natural language processing. He has published several articles in various journals and has participated in numerous conferences. He also serves on the editorial board of several scientific journals.