International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V74I5P102

Volume 74 Issue 5, 10-14, May 2026

Original Article

A Comparative Study of Deep Learning Architectures for Beat-Level Arrhythmia Classification using the MIT-BIH Database

Ceena Mathews

Department of Computer Science, Prajyoti Niketan College, Pudukad, India.

Corresponding Author : ceenamathews@prajyotiniketan.edu.in

Received: 16 March 2026 Revised: 21 April 2026 Accepted: 12 May 2026 Published: 28 May 2026

Abstract *- Cardiovascular diseases continue to be a major cause of mortality worldwide, and among these, cardiac arrhythmias still remain challenging to diagnose accurately. Electrocardiogram (ECG) analysis is a non-invasive tool for the detection of arrhythmias. Manual interpretation of ECG signals can be time-consuming and often varies between clinicians. In this work, six deep learning architectures, ResNet1D-34, InceptionTime, transformer, attention-BiLSTM, attention-aware pooling, and convolution-enhanced transformer, are implemented for beat-level arrhythmia classification using the MIT-BIH arrhythmia database under AAMI class grouping (Normal, Superventricular, Ventricular, Other). All these models are evaluated using macro F1-score to better account for class imbalance, and class-wise evaluation is also assessed to better predict the model performance. The results show that convolution-based models tend to perform more consistently and achieve better balance across classes, while attention-based models struggle with classes that have similar waveform patterns, particularly supraventricular beats.*

Keywords *- Attention, Convolution, ECG beats, InceptionTime, LSTM, Transformer.*

1. Introduction

Cardiovascular diseases remain a leading cause of mortality worldwide. Timely and accurate diagnosis of cardiac arrhythmias is complicated. It can be detected non-invasively through Electrocardiogram (ECG) signals; however, manual interpretation is time-consuming and highly dependent on clinical expertise, often resulting in inter-observer variability. These limitations resulted in the automated ECG analysis systems, which are based on advanced computational techniques.

According to the AAMI standard, ECG beats are categorised into five classes: Normal (N), Supraventricular (S), Ventricular (V), Fusion (F), and other (Q). Among these classes, the fusion class is normally excluded from studies due to its limited representation in the databases.

Initially, arrhythmia classification relied on handcrafted feature extraction combined with traditional machine learning algorithms. Although these methods achieved comparable results, their performance strongly depended on feature engineering and often failed to generalise across datasets. With the advent of deep learning, the models can automatically learn hierarchical feature representations directly from raw ECG signals. Recent research showed that deep neural networks can achieve cardiologist-level performance on large-scale ECG classification tasks [1]. [2][3] highlight that commonly used datasets, including MIT-BIH, exhibit severe class imbalance, with fusion, supraventricular, and other beats having a minimum amount of data representation. Models trained using such an imbalanced dataset show low recall and precision for minority classes while achieving high overall accuracy. Additionally, these studies show that morphological similarity between certain classes, particularly normal and supraventricular beats, causes inconsistent misclassification, even with deep architectures. Preprocessing steps such as filtering, R-peak alignment, and beat segmentation vary significantly between works, which greatly influence model performance [2]. Furthermore, different train–test split strategies (inter-patient vs. intra-patient) can lead to inconsistent optimistic results [3]. Also, evaluation metrics differ, with some studies relying primarily on overall accuracy instead of class-wise metrics, making it difficult to assess performance on clinically important minority classes.

To address these limitations, in this work, six different deep learning architectures are implemented and evaluated for beat-level arrhythmia classification using the MIT-BIH Arrhythmia dataset. The six different architectures considered in this study are ResNet, InceptionTime, transformer-based networks, attention-based BiLSTM models, attention-aware pooling, and a convolution-enhanced transformer. All these models use the same preprocessing pipeline and experimental setup to ensure unbiased comparison. Performance of these models is assessed using class-wise precision, recall, and F1-score, with macro-F1 adopted as the primary evaluation metric to account for class imbalance, particularly for minority arrhythmia classes.

2. Related Work

Conventional machine learning approaches used with ECG classification leverage signal processing techniques combined with handcrafted feature extraction. Although these methods derived better results, they did not work well across different datasets due to feature engineering. Recently, deep learning has gained popularity due to automated feature engineering, the availability of large datasets, and computational resources. Convolutional Neural Networks (CNNs) have emerged as a better choice due to their ability to extract local morphological structures in ECG recordings. Ribeiro et al. [1] signify that Deep Neural Networks trained on large datasets can achieve better performance. Ismail Fawaz et al. [4] used multi-scale convolutional filters with the InceptionTime model to deal with the class imbalance, thus increasing the performance in time-series tasks.

Recurrent models such as Long Short-Term Memory (LSTM) networks have been explored to account for temporal dependencies in ECG waveforms. Moreover, hybrid CNN–LSTM models are found to be more effective than either of the approaches since they use local feature extraction along with time-series modelling [5][6]. To focus on more important and relevant sections of the signal, attention mechanisms are integrated with these architectures. Transformer-based architectures such as ECGformer [7] and ECGTransForm [8] show that they can effectively model long-range dependencies in ECG signals, thus enhancing classification performance. Furthermore, to capture the global context of the ECG recordings along with local morphological feature extraction, transformer models that utilise self-attention have been used for ECG classification, resulting in superior performance [9][10]. Recent works have increasingly moved toward combining multiple architectural ideas, such as convolutional, recurrent, and attention-based components, into unified models. These hybrid designs aim to improve robustness and generalisation, especially in more complex classification settings [2]. Despite these advances, there are several challenges in the accurate and timely diagnosis of cardiac arrhythmia. Foremost is the class imbalance seen widely in datasets such as the MIT-BIH arrhythmia database. Although the overall accuracy of the ECG classification models is high, it is necessary to understand better how effectively they classify each arrhythmia category.

3. Materials and Methods

The MIT-BIH Arrhythmia database consists of 48 annotated half-hour ECG recordings sampled at 360 Hz, collected from 47 patients aged between 32 and 89 years [11]. From these recordings, individual heartbeats were extracted around annotated R-peaks and categorized according to the AAMI standard into four classes: Normal (N), Supraventricular (S), Ventricular (V), and Other (Q).

Each heartbeat was segmented using a fixed window of 216 samples centred on the R-peak, which is consistent with commonly adopted beat-level classification setups. To reduce variations in signal amplitude across different patients and recordings, each segment was normalised using per-beat z-score normalisation. To address the issue of class imbalance, a class-weighted loss function is used during model training. In this study, six deep learning architectures were selected for comparative evaluation, representing a range of modelling approaches for time-series data.

3.1. ResNet-34

ResNet-34 was adapted for one-dimensional ECG signals by replacing standard 2D convolutions with 1D operations. The network is composed of 34 layers organised into residual blocks, where shortcut connections allow the input of each block to be directly added to its output. The residual blocks help to stabilise training and address the vanishing gradient problem. To capture low-level features, it uses an initial convolutional layer and pooling operation, followed by successive residual connections that learn abstract time-series patterns. In the last stage, global average pooling and a fully connected layer are used for ECG classification.

3.2. InceptionTime

To capture patterns at multiple temporal scales, InceptionTime is used. Multiple 1D convolutions with different kernel sizes are used within each module, which operate in parallel, thereby allowing the model to locate both short and long duration features in the ECG signal. To increase gradient flow, residual connections are included.

3.3. Transformer

It uses an attention mechanism to process sequence data. It leverages self-attention in which each element in the sequence evaluates its relationship with all other elements. This is achieved through query, Key (K), and Value (V) representations. Queries are compared with keys to compute attention scores, which are then used to weight the values, generating context-aware features. To learn different types of relationships concurrently, multi-head attentions are used. Positional encoding is also used since transformers do not inherently capture positional information; positional encoding is added to preserve sequence order. These outputs are then passed through feed-forward layers, along with skip connections and normalisation to stabilise training.

3.4. Attention-based Bidirectional LSTM

Here, bidirectional LSTM networks are combined with an attention mechanism. In order to capture contextual information from past and future time, the sequences are processed in both forward and backward directions. To focus more on sections that are relevant for classification, an attention layer is integrated, which assigns varying importance to different parts of the signal.

3.5. Attention-Aware Pooling

It begins with 1D convolution, batch normalisation, Swish activation, and pooling. To capture local morphological features, residual separable blocks are integrated. Subtle waveform variations, such as negative amplitudes, must be well preserved; therefore, the Swish

activation function is used. For temporal aggregation, attentive statistics pooling is employed, generating attention-weighted mean and standard deviation representations. This allows the model to retain both first and second-order statistical information before passing the features to a dense classification layer. Attentive statistics pooling is applied for temporal aggregation, thereby generating attention-weighted mean and standard deviation representations.

3.6. Convolution- Enhanced Transformer

It is a hybrid architecture that combines convolutional layers with Transformer blocks. Convolutional layers are used to initially extract the local features, which thereby reduces sequence length, and then, in order to capture global dependencies, multi-head self-attention is utilised. Separable convolutions are integrated with feed-forward layers within the transformer to sustain local continuity. To produce a fixed-length representation for classification, attentive statistics pooling is also used in this model.

3.7. Evaluation Metrics

The performance of each of these six deep learning models is evaluated class-wise using metrics such as precision, recall, and F1-score. Macro F1 score is also computed to account for class imbalance since it gives equal importance to all classes.

4. Results and Discussion

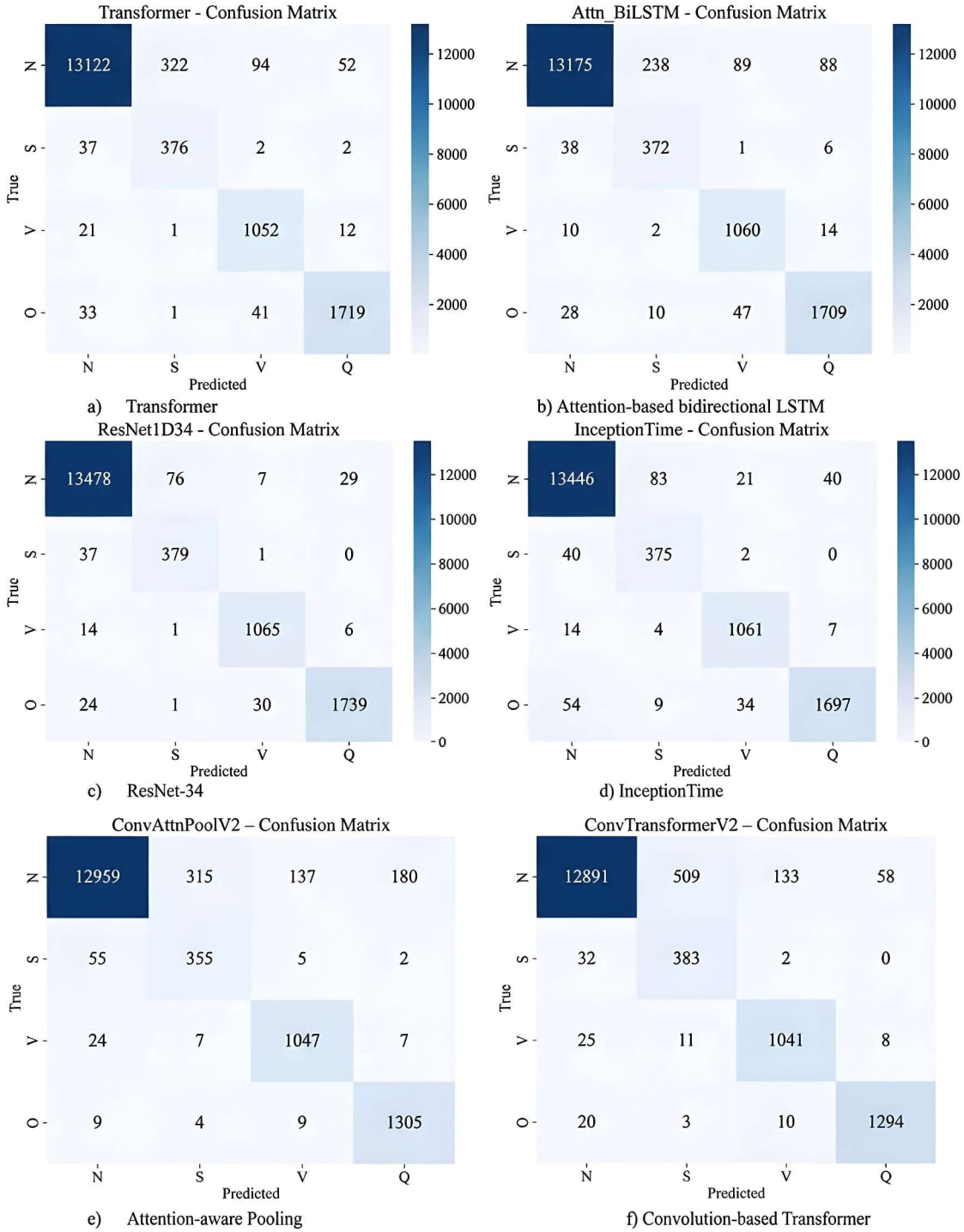
In this study, six deep learning architectures were evaluated for beat-level arrhythmia classification across the four AAMI classes (N, S, V, Q) using the MIT-BIH database. All models were trained under identical conditions using the Adam optimizer with an initial learning rate of 10^{-3}. The results are summarised in Table 1 in terms of per-class precision, recall, and F1-score, while overall accuracy and macro-F1 are reported in Table 2. Across all models, performance on the majority class (N) is consistently high. F1-scores exceeding 0.99 achieved by models such as ResNet1D-34 and InceptionTime indicate that the dominant ECG patterns are learned well. However, all six models produce a low precision for the S class; even the best-performing model, ResNet1D-34, has an F1-score of 0.8673. Furthermore, Transformer-based and attention-driven models produce even lower values, indicating that S and N classes have similar morphological characteristics. However, F1-scores for the V and Q classes in all models are above 0.91, indicating distinct morphological characteristics.

All models achieve higher accuracy, which is above 95. However, as per the macro F1-score, ResNet1D-34 provides the best performance, followed by InceptionTime. Transformer and attention-based models achieve lower macro F1-scores since they produce a lower precision score for minority classes. This indicates that the hybrid models, such as attention-aware pooling and the convolution-enhanced transformer, show moderate improvements. Confusion matrices produced by each of the six deep learning models are presented in Figure 1(a)–(f). It is evident from the confusion matrix of the attention-aware pooling model that the N beats are misclassified as S and Q, indicating that they have similar waveforms with subtle differences. However, V beats has a distinct morphology, which is well indicated in all matrices. The Q beats also have a limited misclassification and hence produce better performance in all cases. Quantitative results and confusion matrices align with each other.

From the results exhibited in both tables and confusion matrices, it can be deduced that convolution-based architectures provide more stable and balanced performance, while attention-based models need further refinement to handle class imbalance and fine-grained morphological variations better. It is also important to note that model performance may vary depending on signal quality, inter-patient variability, and annotation inconsistencies present in the dataset. While this study focuses on beat-level classification, real-world applications require models to generalise across diverse clinical conditions. In addition, variations in preprocessing and segmentation can affect performance. These factors highlight the importance of standardised evaluation to ensure reliable and clinically applicable ECG classification systems.

Table 1. Per-class classification performance of deep learning architectures on the MIT-BIH database under AAMI class grouping

Model	Normal (N)			Superventricular (S)			Ventricular (V)			Other (Q)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
ResNet1D-34	0.9945	0.9918	0.9931	0.8293	0.9089	0.8673	0.9655	0.9807	0.9730	0.9803	0.9693	0.9748
InceptionTime	0.9920	0.9894	0.9907	0.7962	0.8993	0.8446	0.9490	0.9770	0.9628	0.9731	0.9459	0.9593
Transformer	0.9931	0.9656	0.9791	0.5371	0.9017	0.6732	0.8848	0.9687	0.9248	0.9630	0.9582	0.9606
Attention-BiLSTM	0.9943	0.9695	0.9817	0.5981	0.8921	0.7161	0.8855	0.9761	0.9286	0.9406	0.9526	0.9466
Attention-Aware Pooling	0.9863	0.9481	0.9713	0.5221	0.8544	0.6472	0.8712	0.9638	0.9175	0.8654	0.9832	0.9343
Convolution – Enhanced Transformer	0.9888	0.9576	0.9821	0.5171	0.8854	0.6534	0.8743	0.9632	0.9231	0.9714	0.9682	0.9712



a) Transformer b) Attention-based bidirectional LSTM

c) ResNet-34 d) InceptionTime

e) Attention-aware Pooling f) Convolution-based Transformer

Fig. 1 Confusion matrices for deep learning models on the MIT-BIH Arrhythmia dataset

Table 2. Summarised results of all the evaluated models

Model	Accuracy	Macro F1
ResNet1D-34	98.66%	95.21%
InceptionTime	98.18%	93.94%
Transformer	96.34%	88.45%
Attention BiLSTM	96.62%	89.32%
Attention-aware pooling	95.41%	86.55%
Convolution-enhanced Transformer	96.14%	87.93%

5. Conclusion

This paper presented a comparative evaluation of multiple deep learning architectures for beat-level arrhythmia classification using the MIT-BIH Arrhythmia database under the AAMI class grouping. The results indicate that convolution-based models, particularly ResNet1D-34 and InceptionTime, provide more consistent and robust performance, especially when assessed using macro-F1 in the presence of class imbalance.

Although attention-based and hybrid architectures offer advantages in modelling global dependencies, their performance appears limited when distinguishing subtle morphological variations without strong convolutional feature extraction. Analysis of the confusion matrices further shows that most classification errors occur between morphologically similar classes, particularly between Normal and Supraventricular beats.

Future work will focus on extending the evaluation to patient-wise splits, performing cross-dataset validation, and exploring improved strategies for handling minority classes. These directions are expected to enhance both the generalisability and clinical reliability of automated ECG classification systems.

References

[1] Antônio H. Ribeiro *et al.*, "Automatic Diagnosis of the 12-Lead ECG Using a Deep Neural Network," *Nature Communications*, vol. 11, no. 1, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Shenda Hong *et al.*, "Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: A Systematic Review," *Computers in Biology and Medicine*, vol. 122, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yaqoob Ansari *et al.*, "Deep Learning for ECG Arrhythmia Detection and Classification: An Overview of Progress for Period 2017–2023," *Frontiers in Physiology*, vol. 14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Hassan Ismail Fawaz *et al.*, "InceptionTime: Finding AlexNet for Time Series Classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936-1962, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Shahab Ul Hassan et al., "Classification of Cardiac Arrhythmia Using a Convolutional Neural Network and Bi-Directional Long Short-Term Memory," *Digital Health*, vol. 8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mingfeng Jiang et al., "HADLN: Hybrid Attention-Based Deep Learning Network for Automated Arrhythmia Classification," *Frontiers in Physiology*, vol. 12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Taymaz Akan, Sait Alp, and Mohammad Alfrad Nobel Bhuiyan, "ECGformer: Leveraging Transformer for ECG Heartbeat Arrhythmia Classification," *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, pp. 1412-1417, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hany El-Ghaish, and Emadeldeen Eldele, "ECGTransForm: Empowering Adaptive ECG Arrhythmia Classification Framework with Bidirectional Transformer," *Biomedical Signal Processing and Control*, vol. 89, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rui Hu, Jie Chen, and Li Zhou, "A Transformer-Based Deep Neural Network for Arrhythmia Detection Using Continuous ECG Signals," *Computers in Biology and Medicine*, vol. 144, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Donghyeon Kim *et al.*, "A Novel Hybrid CNN-Transformer Model for Arrhythmia Detection without R-Peak Identification using Stockwell Transform," *Scientific Reports*, vol. 15, no. 1, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] G.B. Moody, and R.G. Mark, "The Impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001. [CrossRef] [Google Scholar] [Publisher Link]