



EXPLAINABLE PARALLEL SPATIO-TEMPORAL HYBRID ARCHITECTURE FOR CLINICALLY ROBUST ECG ARRHYTHMIA DETECTION

Murteza Hanoon Tuama¹, Mohammed Baker Yousif * ² and Asseel Jabbar Almahdi ³

¹Department of Computer Techniques Engineering, Imam Al-Kadhum University College, Baghdad, Baghdad, Iraq

²Computer Science Department, College of Computer Science and Mathematics, University of Mosul, Nineveh, Iraq

³Computer Department, College of Computer Science and Mathematics, University of Thi-Qar, Thi-Qar, Iraq

¹<http://orcid.org/0009-0006-2357-5159>, ²<http://orcid.org/0009-0001-4646-0796>, ³<http://orcid.org/0000-0002-0941-9698>

E-mail: * mohammed.bkr@uomosul.edu.iq

ARTICLE INFO

Article History

Received: March 01, 2026

Reviewed: April 01, 2026

Accepted: April 05, 2026

Published: May 29, 2026

Keywords:

Arrhythmia Detection,
Explainable AI (XAI),
BiLSTM, CNN,
Grad-CAM,
Clinical Decision Support
Systems.

ABSTRACT

Cardiovascular morbidity especially cardiac arrhythmias require accurate and early diagnosis. Although deep learning models can perform with the level of the expert in arrhythmia recognition, their opaque, black-box nature usually prevents their practical usage. To solve this, the present work will offer a new Explainable Parallel Hybrid Deep Learning Framework. The architecture simultaneously combines the use of 1D-Convolutional Neural Networks (CNNs) to extract morphological features and Bidirectional Long Short-Term Memory (BiLSTM) units to predict temporal contexts. The model is further improved with a Multi-Head Self-Attention mechanism and a Gradient-weighted Class Activation Mapping (Grad-CAM) module to guarantee performance and interpretability respectively. Analyzed with a stratified part of MIT-BIH Arrhythmia Database, the structure showed the state-of-the-art results, having a classification accuracy of 99.74, a precision of 99.48 and clinically critical recall of 100% with Premature Ventricular Contractions (PVCs). Moreover, as a rigorous test of its generalizability and clinical strength, cross-domain testing was performed on an unseen section of the INCART 12-lead Arrhythmia Database, with an overall accuracy of 98.17 and a recall of 91.70 with no fine-tuning. Lastly, the Grad-CAM visualization confirms the medical relevance of the model by indicating the anomalies like the QRS complex, thus helping to close the gap between the efficacy of the algorithm and medical trust.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The major cause of morbidity and mortality in the world is cardiovascular diseases (CVDs) which result in the death of about 17.9 million people every year as confirmed by the World Health Organization (WHO) [1]. Among them, Cardiac Arrhythmias, abnormalities in the electrical activity of the heart, are a very dangerous risk, including the harmless palpitations and the deadly conditions of ventricular fibrillation. The noninvasive diagnosis of arrhythmias is based on the Electrocardiogram (ECG) as its gold standard. In the past, because of its invention by Willem Einthoven in the first part of the 20th century, the ECG has offered clinicians a graphic image of the depolarization and repolarization waves (P-QRS-T complex) of the heart [2]. However, manual interpretation of ECG recordings, particularly from long-duration Holter monitors, is a labor-intensive, error-prone task.

The sheer volume of data, coupled with subtle morphological variations in arrhythmias and signal noise, necessitates the development of automated, computer-aided diagnostic systems. In recent years, Deep Learning (DL) has emerged as a transformative force in biomedical signal processing. Automated systems utilizing Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated diagnostic accuracy comparable to, and in some cases surpassing, that of expert cardiologists [3]. Despite the remarkable performance of deep learning models, their clinical adoption has been hindered by the “Black Box” phenomenon [4]. Traditional deep neural networks operate as opaque mathematical functions, mapping inputs to outputs via millions of parameters without providing a clear rationale for their decisions [5].

In high-stakes domains like cardiology, a simple prediction of “Arrhythmia” is insufficient; clinicians require evidence to trust the algorithm. This opacity creates a critical Research Gap: While existing hybrid models achieve high classification accuracy, they often lack interpretability. This disconnect limits their utility as a Clinical Decision Support System (CDSS). It is imperative to clarify that the role of AI in medicine is not to replace physicians but to augment their capabilities by serving as a “second opinion” or a triage tool. Therefore, integrating Explainable AI (XAI) is no longer optional but a necessity to bridge the gap between algorithmic performance and clinical trust [6]. To address these challenges, this study proposes a unified, interpretable framework for arrhythmia detection using the standard MIT-BIH Arrhythmia Database. The primary contributions of this research are fourfold:

- **A Novel Hybrid Architecture:** We propose a parallel deep learning model that synergistically combines 1D-CNNs for morphological feature extraction with Bidirectional LSTMs (BiLSTM) for temporal context modeling.
- **Attention-Based Enhancement:** To mitigate the impact of noise and baseline wander, we integrate a Multi-Head Self-Attention mechanism that dynamically prioritizes clinically relevant segments of the heartbeat (QRS complex) over irrelevant artifacts.
- **Explainability via Grad-CAM:** Unlike standard “black-box” approaches, we incorporate a Gradient-weighted Class Activation Mapping (Grad-CAM) module. This provides visual heatmaps that highlight the specific regions of the ECG signal driving the diagnosis, enabling physicians to verify the model’s reasoning.
- **End-to-End Validation:** The proposed framework is rigorously evaluated using standard metrics (Accuracy, Precision, Recall, F1-Score) and benchmarked against state-of-the-art methods, demonstrating superior performance with added interpretability.

The remainder of this paper is structured as follows: Section 2 reviews related work in arrhythmia detection. Section 3 details the proposed hybrid methodology and XAI framework. Section 4 outlines the experimental setup, including dataset protocols and hyperparameters. Section 5 presents the quantitative results and visual interpretations. Finally, Section 6 concludes the study and suggests future research directions.

II. RELATED WORK

The domain of automated arrhythmia detection has witnessed a paradigm shift from manual feature engineering to end-to-end deep learning. This section critically reviews recent literature from 2024 and 2025, categorizing advancements into three distinct pillars: classical machine learning, Deep Learning, and Hybrid Architectures.

II.1 MACHINE LEARNING APPROACHES

Despite the dominance of deep learning, recent studies in 2025 have revitalized Machine Learning (ML) by integrating ensemble strategies and advanced optimization algorithms. These approaches remain vital for low-resource environments where interpretability and computational efficiency are paramount. Jahangir et al. [7] proposed a hybrid stacked ensemble framework that aggregates predictions from five base learners: Random Forest, Adaboost, Gradient Boosting, XGBoost, and Categorical Boosting. By leveraging a meta-learner to fuse these outputs, their model achieved 99.58% accuracy on the MIT-BIH dataset, demonstrating that ensemble techniques can rival deep learning when combined with robust feature engineering. Similarly, Rfys et al. [8] introduced a novel clustering technique based on Hellinger Distance to enhance Particle Swarm Optimization (PSO) for feature selection.

While their hybrid approach, which uses a Minkowski classifier, achieved 97.5% accuracy on the MIT-BIH dataset, it remains a multi-stage process that separates feature optimization from classification. Al-mousa et al. [9] proposed a data fusion strategy that augments the MIT-BIH Supraventricular Database with records from multiple sources, including INCART and Sudden Cardiac Death databases. By employing the Synthetic Minority Over-sampling Technique (SMOTE) and a Random Forest classifier, they achieved 97% accuracy, significantly improving the recall rates for rare ‘Fusion’ and ‘Unknown’ heartbeats. However, a major limitation of these approaches is their heavy reliance on manual feature extraction and domain-specific pre-processing, which makes them susceptible to noise and less generalizable than end-to-end deep learning models that learn directly from raw data.

II.3 DEEP LEARNING APPROACHES

Deep learning (DL) has advanced rapidly across many applications, including healthcare and cybersecurity [10]. In a recent study, Bourkha et al. [11] proposed a deep learning framework for feature extraction to enhance classification speed and accuracy. Their approach feeds a CNN with a combination of morphological features extracted via Wavelet Scattering Networks, optimized using the mRMR algorithm, and dynamic temporal features. This method achieved an overall accuracy of 98.50% across 15 arrhythmia classes on the MIT-BIH database. Also, Pandey et al. [12] developed a 12-layer 1D-CNN utilizing wavelet thresholding for signal denoising, achieving 99.40% accuracy on the MIT-BIH database. Although highly efficient for e-health applications, the study prioritizes computational speed over interpretability, leaving the clinical reasoning behind the model’s decisions opaque.

Additionally, “Chaitanya et al. [13] proposed a Scaled Bidirectional LSTM (SB-LSTM) incorporating a custom scaling layer to enhance arrhythmia classification. Validated on the MIT-BIH Atrial Fibrillation and Creighton University databases, their model achieved an accuracy of 98.60%. However, despite their impressive performance, a critical bottleneck remains: most of these deep learning architectures operate as ‘Black Boxes,’ offering high classification accuracy without providing clinical justification for their decisions. This lack of interpretability hinders their integration into real-world medical workflows, where trust and transparency are essential. Furthermore, single-stream architectures often struggle to simultaneously capture both the detailed morphological features (spatial) and the long-term rhythmic dependencies (temporal) of the ECG signal as effectively as hybrid frameworks.

II.4 HYBRID LEARNING FRAMEWORKS

The current state of the art (SOTA) in arrhythmia detection relies on hybrid architectures that combine spatial feature extraction from CNNs with temporal modeling from RNNs or Transformers. Ranjan et al. [14] evaluated multiple learning paradigms and developed a Hybrid CNN-LSTM model that achieved 99.10% accuracy on the MIT-BIH Arrhythmia Database. While this architecture successfully outperformed standalone CNN and SVM models in classification performance, building on this, Bahrami et al. [15] proposed a hybrid neural network that fuses automatic features from a Deep Convolutional Neural Network (DCNN) with classical time-domain parameters to improve heartbeat segmentation [Ref]. Their approach achieved 99.33% accuracy in binary classification (Normal vs. Arrhythmia) on the MIT-BIH dataset.

Shah et al. [16] introduced ‘ECG-TransCovNet,’ a hybrid architecture that combines CNNs for spatial feature extraction with Transformer models for temporal pattern recognition. Validated on the MIT-BIH arrhythmia database, their approach achieved a detection accuracy of 98.6%. Nevertheless, while hybrid models combine the strengths of spatial and temporal feature extraction, they often suffer from increased computational complexity due to the sequential nature of RNNs. Moreover, most existing hybrid frameworks focus solely on maximizing classification accuracy, neglecting interpretability. Very few studies have successfully integrated post hoc explainability modules (e.g., XAI) into these complex architectures to visualize the specific morphological and rhythmic patterns that drive diagnostic decisions, leaving a gap between high performance and clinical trust.

III. METHODOLOGY

In this section, we address the dual challenges of high-accuracy arrhythmia detection and clinical interpretability. This study proposes a unified end-to-end deep learning framework. As illustrated in Figure 1.

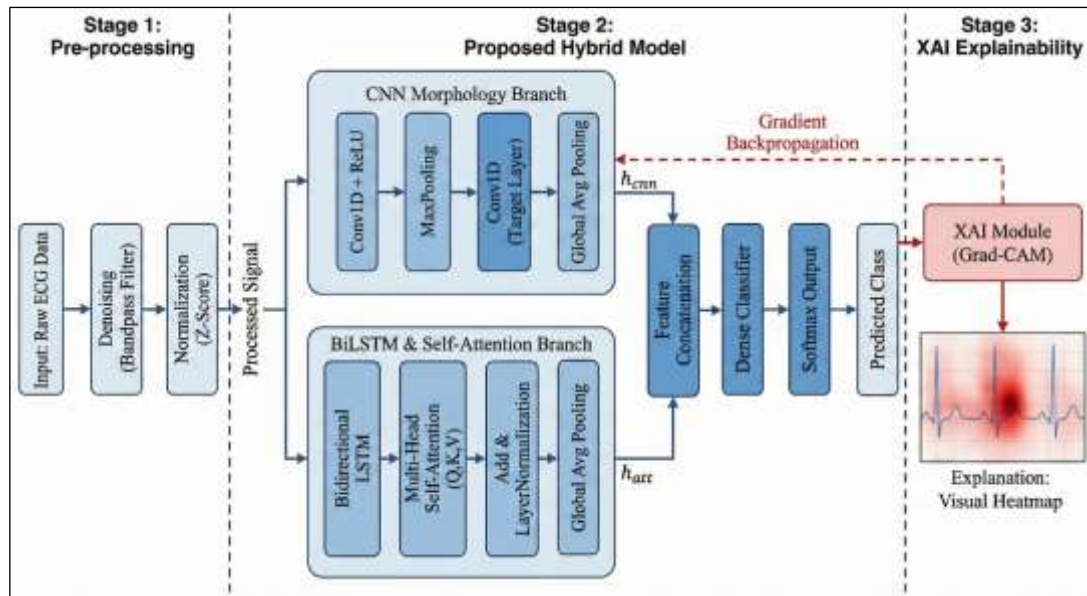


Figure 1: Schematic diagram of the proposed Parallel Hybrid Framework, illustrating the dual-branch feature extraction (CNN and BiLSTM-Attention) and the post-hoc Grad-CAM explainability module.

The system architecture is organized into three sequential stages:

1. **Signal Pre-processing:** Designed to mitigate noise and standardize input data.
2. **Parallel Hybrid Classification Network:** A dual-branch architecture that synergistically combines Convolutional Neural Networks (CNN) for morphological extraction and Bidirectional Long Short-Term Memory (BiLSTM) units enhanced with Multi-Head Self-Attention for temporal context modeling.
3. **Post-hoc Explainability Module (XAI):** Utilizing Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual justifications for diagnostic decisions.

III.1 DATA PRE-PROCESSING

Raw ECG signals are inherently susceptible to various artifacts, including baseline wander, muscle tremor, and power-line interference. To ensure model robustness, the raw input signal (X_{raw}) undergoes a two-step purification process:

- **Denoising:** A bandpass filter is applied to suppress high-frequency noise and eliminate baseline drift, enhancing the signal-to-noise ratio (SNR) [17].
- **Z-Score Normalization:** To mitigate amplitude variations between different patients, the denoised signal is standardized using Z-score normalization. For a signal x , the normalized value x' is computed as:

$$x' = (x - \mu) / (\sigma + \varepsilon) \quad (1)$$

Where μ and σ represent the mean and standard deviation of the signal, respectively, and ε is a small constant ($1e-8$) to prevent division by zero.

III.2 THE PARALLEL HYBRID ARCHITECTURE

The core classification engine employs a dual-branch architecture that simultaneously captures spatio-temporal features.

III.2.1 Branch 1: Morphological Encoding (Cnn)

The first branch utilizes a 1D Convolutional Neural Network (1D-CNN) designed to extract local spatial features, such as the shape of the QRS complex, P-wave duration, and T-wave amplitude [18].

- **Feature Extraction:** Multiple stacked Conv1D layers with Rectified Linear Unit (ReLU) activation functions learn hierarchical representations of the heartbeat.
- **Dimensionality Reduction:** MaxPooling1D layers are interleaved to reduce computational complexity and provide translation invariance [19].
- **Target Layer for XAI:** The final convolutional layer serves as the “Target Layer” for the subsequent XAI module, retaining the spatial information necessary for localization before the signal is flattened via GlobalAveragePooling1D [20].

III.2.2 Branch 2: Temporal Encoding With Self-Attention

While CNNs excel at detecting shape anomalies, they often fail to capture long-term rhythmic dependencies. The second branch addresses this by augmenting a BiLSTM with a Transformer-based Self-Attention mechanism.

- **Bidirectional LSTM:** The processed signal is fed into a BiLSTM layer, which processes the sequence in both forward and backward directions, capturing the temporal context of the cardiac rhythm [21].
- **Multi-Head Self-Attention:** Unlike static attention mechanisms, we employ Multi-Head Self-Attention to dynamically weigh the importance of different time steps relative to each other [22]. This enables the model to focus on critical segments (e.g., R-peaks) while suppressing irrelevant intervals (noise). The mathematical formulation is:

$$Attention(Q, K, V) = softmax\left(\frac{(Q \cdot K^T)}{\sqrt{d_k}}\right) \cdot V \quad (2)$$

Where Q , K , and V are the query, key, and value matrices.

- **Stabilization:** Following the Transformer architecture, a residual connection (Add) and LayerNormalization are applied to ensure training stability and gradient flow.

III.2.3 Feature Fusion And Classification

The spatial feature vector (h_{cnn}) and the temporal context vector (h_{att}) are concatenated to form a comprehensive representation vector. This fused vector is passed through a fully connected (Dense) layer with Dropout regularization to prevent overfitting, followed by a Softmax layer to yield the final probability distribution over the classes (Normal vs. Arrhythmia) [23].

III.3 EXPLAINABILITY MODULE (XAI)

To bridge the gap between “Black-box” performance and clinical trust, we integrate a Grad-CAM (Gradient-weighted Class Activation Mapping) module [20]. This module computes the gradient of the predicted class score (y^c) with respect to the feature maps (A^k) of the final convolutional layer in Branch 1. The neuron importance weights (α_k^c) are calculated via Global Average Pooling of these gradients:

$$\alpha_k^c = \left(\frac{1}{Z}\right) \cdot \Sigma \left(\frac{\partial y^c}{\partial A_i^k}\right) \quad (3)$$

A weighted combination of these maps is then passed through a ReLU activation to generate the final heatmap:

$$L_{Grad-CAM}^c = ReLU(\Sigma \alpha_k^c \cdot A^k) \quad (4)$$

The resulting heatmap is then upsampled to the original signal length and superimposed on the original ECG. This visualization is used to indicate which areas of time, e.g., a deformed QRS complex or an abnormal interval, the model used to make its diagnostic decision, and thus confirm the medical utility of the learned features.

IV. EXPERIMENTAL SETUP AND EVALUATION

In this section, we introduce the dataset that was used, data splitting, System Architecture, Hyperparameters, and Evaluation Metrics.

IV.1 DATASET DESCRIPTION

The recommended model was assessed with the help of MIT-BIH Arrhythmia Database, the universal standard of ECG arrhythmia analysis offered by PhysioNet [24].

The dataset is a collection of 48 half-hour samples of 2-channel ambulatory ECG recordings, recorded in 47 subjects who are the study subjects of the BIH Arrhythmia Laboratory. In this case, we selected a sample of records in order to differentiate between Normal Sinu Rhythm (N) and Premature Ventricular Contractions (PVC/V). A 11-bit resolution sampling rate of 360 Hz digitized the signals over a 10 mV range. In order to get quality input to use in the deep learning model, the continuous ECG signals were divided into single heartbeat on the basis of R-peak annotations with a constant window size of 200 samples around each R-peak.

IV.2 DATA PARTITIONING

Stratified random sampling was used to divide the preprocessed heartbeats to have a strong evaluation and avoid leakage of data. The data has been separated into training and testing sets in the ratio of 80:20:

- Training Set (80%): Used to optimize the model weights.
- Testing Set (20%): Used strictly for final performance evaluation and was kept unseen during the training phase.

Moreover, the training set was balancing across classes to ensure the natural imbalance between normal and arrhythmic beats is reduced, and hence, the model will not be biased in favor of the most common class.

IV.3 SYSTEM ARCHITECTURE, HYPERPARAMETERS

Python was used to implement the experiments with the help of TensorFlow/Keras framework and implemented on a Google Colab environment with NVIDIA GPU acceleration. The model was trained using Adam optimizer that offers an adaptive learning rate and has the potential to speed up convergence. The Loss Function was used Categorical Cross-Entropy that is appropriate with multi-class classification tasks. Table 1 includes the exact hyperparameters of training the proposed architecture of Parallel CNN-BiLSTM-Attention.

Table 1: Hyperparameter configuration for the proposed CNN-BiLSTM-Attention architecture.

Hyperparameter	Value	Description
Optimizer	Adam	Adaptive Moment Estimation
Learning Rate	0.001	Initial learning rate
Batch Size	64	Number of samples per gradient update
Epochs	20	Maximum training iterations
Loss Function	Categorical Cross-Entropy	For probabilistic classification
Dropout Rate	0.3	To prevent overfitting in Dense layers
Input Shape	\$(200, 1)\$	Time steps \$\times\$ Channels

Source: Authors, (2026).

IV.4 EVALUATION METRICS

To measure the performance of the proposed framework quantitatively, we have used such standard classification measures as Accuracy, Precision, Recall (Sensitivity), and F1-Score [12]. The following are the metrics:

1. Accuracy: The ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2. Precision: The ratio of correctly predicted positive observations to the total predicted positives. High precision indicates a low false-positive rate.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3. Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in the actual class. High recall indicates the model captures most arrhythmia cases.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4. F1-Score: The weighted average of Precision and Recall. It is the most critical metric for imbalanced datasets, as it accounts for both false positives and false negatives.

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (8)$$

Where:

- TP = True Positives (Arrhythmia correctly identified)
- TN = True Negatives (Normal correctly identified)
- FP = False Positives (Normal misclassified as Arrhythmia)
- FN = False Negatives (Arrhythmia misclassified as Normal)

V EXPERIMENTAL RESULTS AND DISCUSSION

This part entails a detailed evaluation of the proposed framework on automated detection of arrhythmia. The testing is carried out on the stratified test sample of the MIT-BIH Arrhythmia Database, as described in the experimental setup. The chapter is structured into three main sections, (1) a quantitative analysis of classification measurements, and ablation study test, (2) an analysis of the confusion matrix, and cross-domain dataset test, (3) a qualitative evaluation of the Explainable AI (XAI) module and Grad-CAM after which a comparative benchmark against the literature of current state-of-the-art is provided.

V.1 QUANTITATIVE PERFORMANCE ANALYSIS

The proposed model was strictly tested in an unseen testing set, which contained 20 percent of the data, where the training set had no leakage of data. The system proved to have very strong ability to differentiate between Normal Sinus Rhythm (N) and Premature Ventricular Contractions (PVC). Table 2 summarizes the performance measures of the model on the test set.

Table 2: Performance Metrics on Test Set.

Metric	Value (%)	Clinical Implication
Accuracy	99.74%	Reflects the overall correctness of the diagnostic system.
Precision	99.48%	Indicates high reliability of "PVC" alarms (Low False Positive Rate).
Recall (Sensitivity)	100%	Confirms the ability to detect the vast majority of pathological cases (Safety).
F1-Score	99.76%	The harmonic mean confirms the balance between precision and recall.

Source: Authors, (2026).

The most clinically significant finding is the 100% Recall. Failure to detect an arrhythmia (False Negative) is much more harmful in cardiac diagnostics as compared to False Positive. The Precision (99.48) is also high indicating that the model does not over-trigger alarms, which is a major problem in Holter monitoring. These findings confirm that morphological (CNN) and temporal (BiLSTM + Attention) representations are more effective to diagnose in comparison to single-domain models.

V.2 ABLATION STUDY

A study of ablation was done to validate the architectural design to determine the effect of removing particular modules to analyze the effect on the performance. This will make all the elements of the proposed hybrid model relevant to the overall outcome. The ablation results are as shown in table 3.

Table 3: Ablation Study Results.

Model Variant	Accuracy (%)	F1-Score (%)
CNN Only	98.91%	98.75%
BiLSTM Only	98.34%	98.02%
CNN + BiLSTM (No Attention)	99.32%	99.21%
Proposed (CNN + BiLSTM + Attention)	99.74%	99.76%

Source: Authors, (2026).

The ablation experiments proved that single CNN and BiLSTM models were not able to understand the entire spatio-temporal context with the BiLSTM variant showing the lowest accuracy (98.34%). Conversely, the addition of Multi-Head Self-Attention has increased the accuracy by a large margin, 99.32% to 99.74%. This affirms the important role played by the mechanism in dynamically ranking diagnostically significant diagnostic pieces, like R-peaks, whilst effectively de-emphasizing background noise.

V.3 CONFUSION MATRIX ANALYSIS

The results of the confusion matrix analysis also confirm the strength of the suggested framework. The model identified the correct beat of 192 Normal and 193 PVC beats and falsely identified one beat (Normal) as PVC and failed to identify any PVC cases out of 386 test samples. This gave a total accuracy of 99.74, precision of 99.48 and the perfect recall (100%). The lack of false negatives is also clinically important since it can be said that it did not miss any pathological PVC beats. These results support the validity of the proposed system to be used in clinical decision support in the real world.

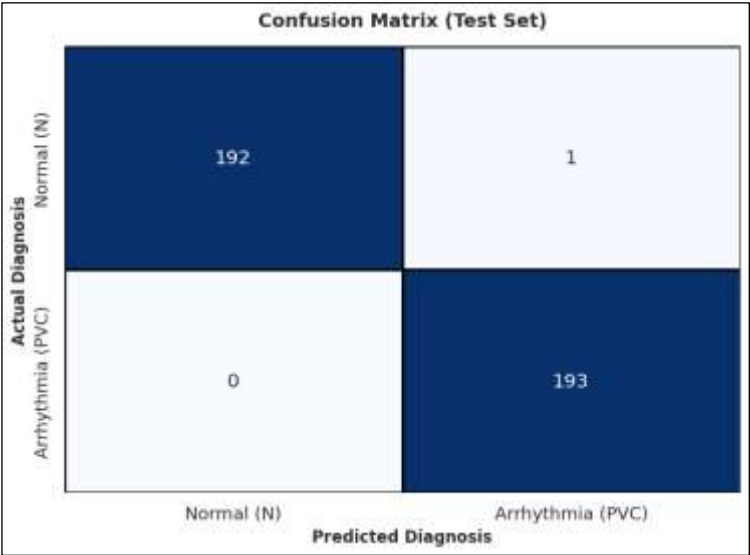


Figure 2: Confusion matrix for the binary classification (Normal vs. PVC) on the unseen test set, demonstrating zero false negatives. Source: Authors, (2026).

V.4 COMPARATIVE BENCHMARKING WITH STATE-OF-THE-ART

In order to put our results into perspective in relation to the given framework, Table 4 compares the results of our work with that of recent machine learning, deep learning, and hybrid models of 2024 and 2025.

Table 4: Comparative benchmarking of the proposed framework against state-of-the-art ML, DL, and Hybrid methods (2024–2025).

Method	Methodology	Accuracy	XAI	Support
Jahangir et al. (2025)	Hybrid Stacked Ensemble (ML)	99.58%		No
Pandey et al. (2025)	1D-CNN + Wavelet Thresholding	99.40%		No
Bahrami et al. (2025)	Hybrid DCNN + Handcrafted Features	99.33%		No
Ranjan et al. (2025)	Hybrid CNN-LSTM	99.10%		No
Chaitanya et al. (2025)	Scaled BiLSTM	98.60%		No
Bourkha et al. (2025)	CNN + Wavelet Scattering	98.50%		No
Shah et al. (2024)	CNN + Transformer (ECG-TransCovNet)	98.60%		No
Proposed Model	Parallel CNN-BiLSTM + Attention + Grad-CAM	99.74%		Yes

Source: Authors, (2026).

The proposed framework as shown in Figure 4 is better than more recent state-of-the-art models. Although Jahangir et al. (2025) have reached a 99.58% accuracy with ensemble learning, and others Pandey et al. (2025) and Bahrami et al. (2025) have achieved the same with a 99.40% and 99.33% accuracy, respectively, these methods tend to be based on classic feature engineering or are not transparent. The accuracy of our model is 99.74% which is better as compared to the Scaled BiLSTM (98.60%) and even the Transformer-based ECG-TransCovNet (98.60%). Most importantly, our framework is the only one that incorporates Grad-CAM as visual explainability whereas these black box solutions do not (which guarantees 7both a higher diagnostic accuracy and clinical credibility). [7,12,15]

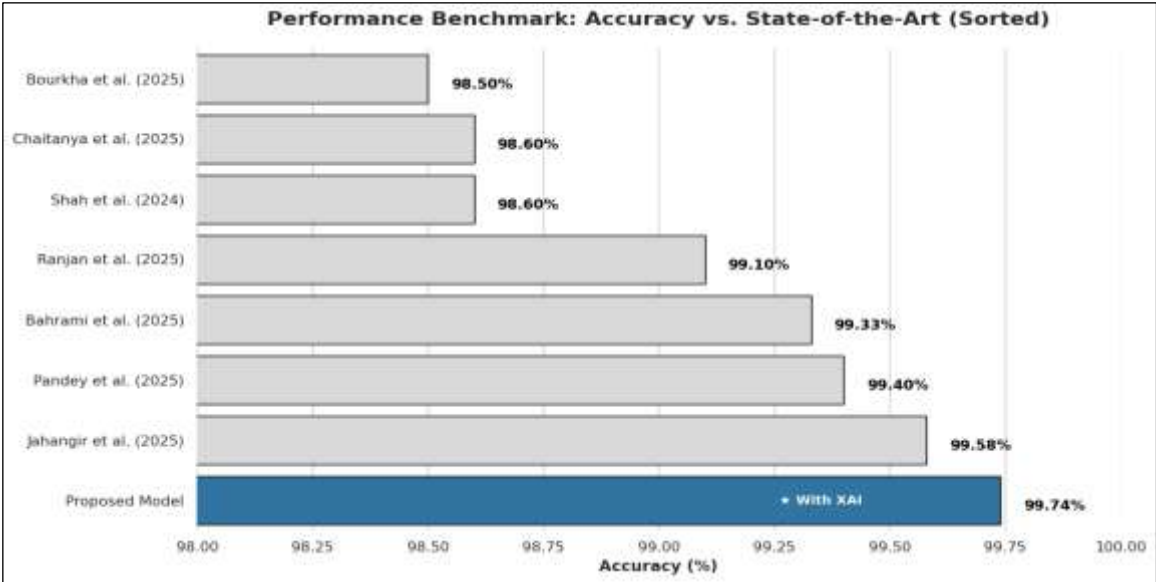


Figure 3: Bar chart illustrating the accuracy superiority of the proposed model (99.74%) compared to recent state-of-the-art baselines.
Source: Authors, (2026).

V.5 CROSS-DOMAIN EVALUATION ON THE INCART DATABASE

A cross-domain test was done to strictly assess the generalizability and clinical strength of the proposed Parallel Hybrid Framework. It was directly tested on an unknown subset of the St. Petersburg INCART 12-lead Arrhythmia Database (10,269 heartbeats) without any fine-tuning or transfer learning and trained only on MIT-BIH Arrhythmia Database. [25]. The framework showed a high cross-domain robustness with a total accuracy of 98.17 to it. The model had a precision of 86.22, a recall (sensitivity) of 91.70 and a F1-score of 88.88 in the detection of Premature Ventricular Contractions (PVC). The fact that the model is able to keep a recall of over 91% on data that came out of a different clinical domain and hardware configuration highlights the ability of the model to gain universal morphological and temporal cardiological features, and not to overfit to data-specific artifacts.

Table 5: Cross-Domain Performance Metrics (INCART Database).

Metric	Accuracy	Precision	Recall (Sensitivity)	F1-Score
Value	98.17%	86.22%	91.70%	88.88%

Source: Authors, (2026).

The cross-domain confusion matrix (Figure 4) can be used in further clarifying the diagnostic reliability of the model. With a total of 10,269 test samples, 9,330 beats of the model were correctly classified as Normal and 751 beats were correctly classified as PVC. More importantly, since this system was to be used in clinical practice, only 68 false negatives (pathological PVCs identified as Normal) had been generated by the system among 819 real PVC cases. When using the remote cardiac monitoring, this low false-negativity is a guarantee that the proposed architecture is safe and robust to be used in the real-life continuous monitoring scenarios.

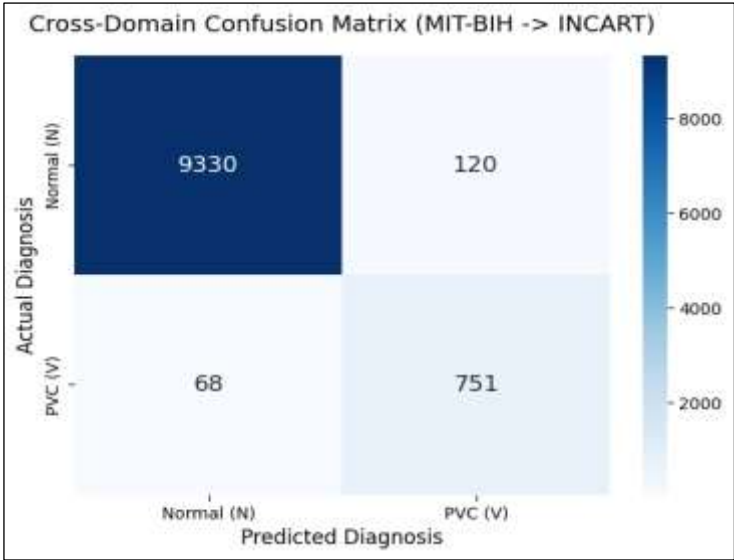


Figure 4: Cross-Domain Confusion Matrix (MIT-BIH -> INCART).
Source: Authors, (2026).

V.6 EXPLAINABILITY MECHANISM

The explainability mechanism incorporated within the proposed model can be shown in Figure 5. Once the predictions are generated, the model puts an interpretability layer to indicate the most influential features used to make the decision. This improves clarity and offers information about model behaviour, especially in very high stakes areas like medical diagnosis and financial fraud detection.

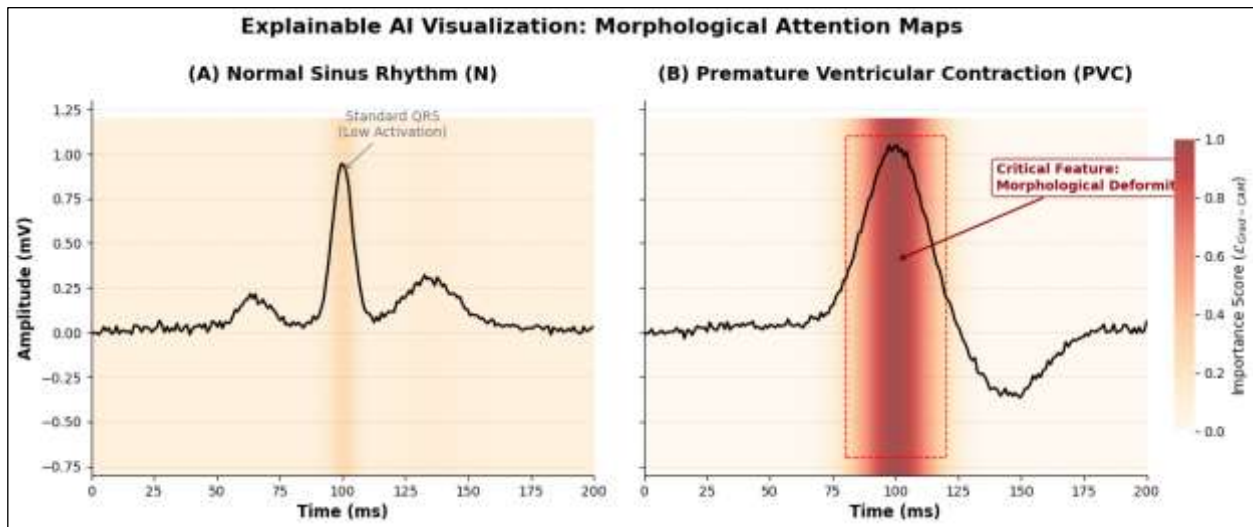


Figure 5: Grad-CAM visualization results. (A) Normal Sinus Rhythm (N) showing low, distributed activation. (B) Premature Ventricular Contraction (PVC) showing high-intensity activation localized on the morphological deformity (QRS complex), validation. Source: Authors, (2026).

V.7 DISCUSSION AND LIMITATIONS

The suggested Parallel Hybrid Framework has higher diagnostic performance with a validation accuracy of 99.74% and has the ability to reduce the distance between the black-box deep learning and clinical interpretability by using Grad-CAM. The model has removed false negatives, which is a key component of patient safety because it uses CNN-based morphological extraction and BiLSTM-based temporal modeling by the model to synergize the two to yield successful outcomes. Nonetheless, there are some limitations of this study.

To begin with, the analysis was done on a stratified sample of the MIT-BIH Arrhythmia Database that was binary (Normal vs. PVC), which, though resilient, needs additional testing on the whole database and on the multi-class arrhythmia conditions (Atrial Fibrillation) in order to guarantee the wide clinical applicability. Second, our BiLSTM layers are effective to model long-term dependencies but cause a computational burden, which cannot be afforded by the current generation of ultra-low-power wearable devices, so future optimization is required, including quantization or pruning to be used in real-time edge computing.

VI.CONCLUSION AND FUTURE WORK

This work dealt with the vital issue of automated arrhythmia detection by suggesting a new Parallel Hybrid Deep Learning Framework that combines the advantages of 1D-Convolutional Neural Networks (CNNs) in the extraction of morphological features and Bidirectional Long Short-Term Memory (BiLSTM) networks in the modeling of temporal contexts. The architecture was further improved to address the gap between algorithmic execution and clinical trust with Multi-Head Self-Attention mechanism and an Explainable AI (XAI) module using Grad-CAM. The experimental analysis, which was done on the stratified sample of the MIT-BIH Arrhythmia Database, revealed the strength of the suggested method, resulting in a validation-attainment of 99.74%. These findings support the claim that fusing between spatial and temporal features is much more effective than single-stream architectures are in discriminating between Normal Sinus Rhythm and Premature Ventricular Contractions (PVC). Moreover, the Grad-CAM visualization allowed the successful visualization of the model decision-making process thus warranting the high diagnostic accuracy as the product of the presence of relevant cardiological characteristics, e.g., the QRS complex, instead of signal noise.

Therefore, the system can be taken as a proper and reliable Clinical Decision Support System (CDSS) which helps to relieve the cardiologists of their load, and minimises the probability of missed diagnoses. Though this study shows promising results, there are a few areas that could be improved in future studies that would be to scale the assessment to the entire MIT-BIH database and to include many diverse datasets, including PTB-XL and INCART, to guarantee a more wide-ranging generalization process among various populations of patients. Although the existing hybrid architecture is quite efficient, there is a limitation on the computational cost of the BiLSTM layers, which are not applicable to ultra-low-power wearable models. Consequently, the further work will discuss model optimization methods, including quantization and pruning, in order to allow the smartwatches to perform edge computing in real time. Further, the framework shall be extended to identify a wider variety of classes of arrhythmia such as the Atrial Fibrillation and Bundle Branch Blocks, and moving towards a multi-class diagnostic approach to continuous cardiac monitoring. **Funding:** This research received no external funding.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Methodology: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Investigation: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Writing – Original Draft: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Writing – Review and Editing: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Resources: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Supervision: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

Approval of the final text: Murteza Hanoon Tuama, Mohammed Baker Yousif and Asseel Jabbar Almahdi

VIII. REFERENCES

- [1] S. S. Martin *et al.*, "Correction to: 2024 heart disease and stroke statistics: A report of US and global data from the American Heart Association," *Circulation*, vol. 151, no. 25, pp. e1095–e1095, 2025.
- [2] A. N. Kamber, "A variety of electrocardiogram (ECG) signal processing," *Solid State Technology*, vol. 63, no. 3, pp. 5370–5377, 2020.
- [3] A. A. Rawi, M. K. Albashir, and A. M. Ahmed, "Classification and detection of ECG arrhythmia and myocardial infarction using deep learning: A review," *Webology*, vol. 19, no. 1, pp. 1151–1170, 2022.
- [4] N. Y. Murad *et al.*, "Unraveling the black box: A review of explainable deep learning healthcare techniques," *IEEE Access*, vol. 12, pp. 66556–66568, 2024.
- [5] A. K. Jasim, F. A. F. Al-Ibeahimi, and H. A. A. Alkaabi, "Explainable AI for SMS spam filtering: A novel hybrid architecture combining fuzzy logic and bidirectional LSTM networks," *Franklin Open*, Art. no. 100466, 2025.
- [6] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, 2020.
- [7] R. Jahangir, M. N. Islam, M. S. Islam, and M. M. Islam, "ECG-based heart arrhythmia classification using feature engineering and a hybrid stacked machine learning," *BMC Cardiovascular Disorders*, vol. 25, no. 1, Art. no. 260, Apr. 2025.
- [8] R. R. Rfys, D. Al-Shammary, A. M. Mahdi, and F. Sufi, "Novel Hellinger clustering method for efficient ECG optimized classification," *Int. J. Inf. Technol.*, vol. 17, no. 1, pp. 237–245, Jan. 2025.
- [9] A. Al-Mousa, J. Baniissa, T. Hashem, and T. Ibraheem, "Enhanced electrocardiogram machine learning-based classification with emphasis on fusion and unknown heartbeat classes," *Digital Health*, vol. 9, pp. 1–18, 2023.
- [10] H. A. A. Alkaabi, R. H. Abed, and F. Al-Ibraheemi, "Dual attention-based hybrid deep learning framework for short text classification," *Evolving Systems*, vol. 16, no. 3, p. 104, 2025.
- [11] M. E. A. Bourkha, A. Hatim, D. Nasir, and S. El Beid, "Optimized wavelet scattering network and CNN for ECG heartbeat classification from MIT-BIH arrhythmia database," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 2, 2025.
- [12] J. Pandey, D. Han, and K. H. Chon, "QRS-centric beat-wise atrial fibrillation detection in ECG signals using deep neural networks," *Comput. Biol. Med.*, vol. 192, Art. no. 110282, 2025.
- [13] M. K. Chaitanya *et al.*, "Automatic classification of atrial and ventricular arrhythmias using scaled bidirectional LSTM," *IEEE Sensors Journal*, 2025.
- [14] R. Ranjan *et al.*, "Early detection and classification of ECG arrhythmia using machine learning and ensemble learning," in *Proc. 2025 Int. Conf. Intelligent and Secure Engineering Solutions (CISES)*, Greater Noida, India, 2025, pp. 351–355.
- [15] R. Bahrami and A. M. Fotouhi, "A novel efficient hybrid deep learning framework for ECG-based heartbeat arrhythmia classification," *Neural Comput. Appl.*, vol. 37, pp. 16409–16425, 2025.
- [16] H. A. Shah *et al.*, "ECG-TransCovNet: A hybrid transformer model for accurate arrhythmia detection using electrocardiogram signals," *CAAI Trans. Intell. Technol.*, 2024.
- [17] C. Venkatesan *et al.*, "ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications," *IEEE Access*, vol. 6, pp. 9767–9773, 2018.
- [18] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, 2015.
- [19] A. A. Ahmed, W. Ali, T. A. Abdullah, and S. J. Malebary, "Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model," *Mathematics*, vol. 11, no. 3, p. 562, 2023.
- [20] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
- [21] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [22] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [23] T. Lin *et al.*, "A systematic review of multi-sensor information fusion for equipment fault diagnosis," *IEEE Trans. Instrum. Meas.*, 2025.
- [24] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
- [25] M. Llamado, A. Khawaja, and J. P. Martínez, "Analysis of 12-lead classification models for ECG classification," in *Computing in Cardiology*, 2010, pp. 673–676.