



A deep hybrid CNN–BiGRU model for inter-patient ECG beat classification

 Ielapaka Sunil
Kumar¹⁺

 Bhukya Suribabu
Naick²

^{1,2}Department of Electrical Electronics and Communication Engineering,
GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.

¹Email: ielapak@gitam.in

²Email: sbhukya@gitam.edu



(+ Corresponding author)

ABSTRACT

Article History

Received: 23 January 2026

Revised: 12 May 2026

Accepted: 26 May 2026

Published: 30 September 2026

Keywords

Arrhythmia detection

CNN–BiGRU

ECG beat classification

Electrocardiogram

Heartbeat classification

INCART

MIT-BIH.

This study aims to develop and validate a robust machine learning-based framework for automated ECG beat classification, with the purpose of supporting rapid arrhythmia detection and improving the efficiency of clinical decision support systems. The work focuses on assessing the effectiveness of hybrid deep learning architectures that can jointly capture local morphological patterns and sequential dependencies in ECG signals. The proposed experimental framework was designed using publicly available MIT-BIH Arrhythmia and AAMICS INCART databases. ECG recordings were pre-processed and augmented through Gaussian noise addition, amplitude reduction, and temporal shifting to improve model generalization under signal variability. Several deep learning models were trained and evaluated, including convolutional and hybrid recurrent architectures, to identify the most effective approach for ECG record analysis. The findings demonstrate that the CNN–BiGRU architecture achieved the best overall performance, with an accuracy of 93.89%, precision of 93.36%, recall of 93.91%, and F1-score of 93.39%. These results indicate that combining convolutional neural networks with bidirectional gated recurrent units improves the representation of both morphological and temporal ECG characteristics. The study provides proof of concept for using hybrid deep learning architectures in automated ECG analysis. The practical implication of this work is that the proposed CNN–BiGRU framework may support faster, more reliable ECG interpretation and has potential applicability in real-time arrhythmia screening, wearable monitoring systems, and clinical decision support environments.

Contribution/Originality: This study is original in its unified evaluation of hybrid CNN–recurrent architectures using harmonized MIT-BIH Arrhythmia and AAMICS INCART datasets with signal augmentation. The CNN–BiGRU model uniquely combines ECG morphological feature extraction and bidirectional temporal modelling, achieving superior classification performance for automated arrhythmia detection.

1. INTRODUCTION

Cardiovascular diseases remain one of the leading causes of morbidity and mortality worldwide, and cardiac arrhythmias are among the major contributors to sudden cardiac death and long-term cardiac complications [1, 2]. The electrocardiogram (ECG) is a non-invasive, low-cost, and widely accessible diagnostic tool for monitoring the electrical activity of the heart, and it plays a central role in early diagnosis, continuous monitoring, and timely clinical intervention [3]. Consequently, the development of accurate automated ECG beat classification systems has become an important research problem in intelligent healthcare. Traditional ECG beat classification has relied on multiple tasks, including QRS detection followed by segmenting beats, elucidating certain temporal and

morphological characteristics, and finally employing a rule-based or other forms of shallow-level machine-learning classification [4-9]. Several of these methods have proven useful in certain situations; however, most of these methods have little or no robustness when faced with inter-patient differences, heterogeneous acquisition protocols, signal degradation, and data class imbalance [9-11]. In addition, methods of acute feature extraction are very labor-intensive, and many of them don't capture even a small number of features that may describe highly complicated events or complex patterns of drug abuse [12]. The introduction of deep neural networks has revolutionized the analysis of ECG recordings, especially when combined with raw ECG signals or signals that have had little pre-processing [3, 13]. Some of the neural networks that have proven most effective are the Convolutional Neural Networks (CNNs), used for hierarchical classification of morphologies descriptive of features of signals like the QRS complex or segments of ECG recordings [14-16]. Almost all recurrent neural networks, like the Long Short-Term Memory (LSTM) and the Gated Recurrent Units (GRU), have shown excellent performance for learning the internal dependencies and rhythm or timing dynamics of the ECG [17-20]. Thus, the use of CNNs and RNNs in a single combination is deemed appropriate for learning the spatial and local morphologies and dynamics of the ECG signal [21, 22].

First, many studies of ECG beat classification based on a single benchmark dataset, most often MIT-BIH, might produce optimistic generalization estimations considering relatively homogeneous conditions [19, 21, 23]. Second, a few impact studies demonstrated robust results on large proprietary or ambulatory ECG datasets, the conditions of which might not be entirely comparable to open beat-level benchmark assessment utilizing standard AAMI heartbeat classes [3]. Third, most studies concentrate on the performance of a single pre-defined architecture, while few works provide a systematic study of independent models such as CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU all configured under the same preprocessing, augmentation, optimization, and evaluation conditions [18, 19, 21, 22]. Therefore, the impact of the model architecture and bidirectionality on ECG beat classification performance remains unclear. Additionally, the analysis of cross-database studies has been scarce despite the fact that different patient cohorts, signal quality, lead configuration, and acquisition protocols provide a more clinical and relevant environment for the assessment of the model robustness [24]. To address these gaps, the primary goal of this study is to examine the ability of hybrid convolutional-recurrent models to provide more robust and general ECG beat classification under varying cross-database conditions, in comparison to a stand-alone CNN. For this purpose, CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU models have been implemented and assessed in an integrated architecture with the harmonized partitions of the MIT-BIH Arrhythmia Database and the AAMICS INCART Database. Custom mixed activation functions of ReLU, Swish, ELU, and Tanh are also employed to further augment the capabilities of the model through improved non-linearity and gradient flow, and generalization of the model [12, 25]. This study looks at how architecture choices affect ECG beat classification in near real-world scenarios. It does this by integrating datasets that have varied retrieval traits and applying a consistent training and testing approach. The suggested framework attempts to decrease the inter-record variances and bolster the resilience of the models against habitual artifacts found in ECG data by normalizing and standardizing, alongside data augmentation practices that include the addition of white noise, adjustment of the amplitude, and shifting in time. Each of the models was subjected to training using the same interdependent parameter and framework and was contrasted against one another in terms of accuracy, precision, recall, and F1 score, given that the incidence of certain types of beats in a dataset often constitutes a relatively small fraction of the total volume of the ECG beat dataset. This approach provides unrestricted modeling practices as it pertains to recurrent models and bidirectional learning in the field of ECG beat classification. The primary aspects of this research paper have been encapsulated as follows:

- The Literature includes the first comprehensive deep learning framework on the MIT-BIH and the AAMICS INCART datasets for ECG beat classification.

- This work employs a synchronized approach using normalization, standardization, data augmentation, and varied activation functions for enhanced consistency in cross-database assessments.
- This work remains one of the few studies exploring independent setups for CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU models in the same interpatient scenario.
- The paper critically analyses the role of temporal modelling and bidirectionality in ECG beat classification and supports the use of sequential context for enhanced prediction.
- The main findings of the paper suggest that the CNN-BiGRU model provides the optimal overall balance for classification performance, consistency, and computational efficiency in the context of responsive ECG monitoring.

The rest of the paper is organised as follows: Section 2 covers ECG classification using deep learning; Section 3 provides the datasets and materials; and Section 4 presents the proposed methodology, architecture, augmentation, and preprocessing. Section 5 presents experimental setup; Section 6 presents and explains the evaluation protocol and the experimental results. Then, Section 7 explains and elaborates on the work, its constraints, and outlines the next steps. Finally, Section 8 presents the study's conclusion.

2. RELATED WORKS

The classification of automated ECG has transitioned from manual feature selection and categorization to advanced machine learning systems. Designers of previous systems created distinctive descriptors for every single heartbeat, such as morphological shape as well as numerous descriptors for traditional machine learning systems, or their neural counterparts relied heavily on the utilization of RR-interval statistics, wavelet coefficients, or independent components for classification [4-9]. In a generic formulation, a beat segment $x_i \in \mathbb{R}^L$ was transformed into a feature vector $\phi(x_i)$, and the predicted class was obtained as:

$$\hat{y}_i = \arg \max_{c \in \{1, \dots, C\}} p(y_i = c \mid \phi(x_i)) \quad (1)$$

Table 1. Structured comparison of representative ECG classification studies.

Study	Dataset	Architecture / Method	# Classes	Evaluation Metrics	Best Accuracy / F1
De Chazal, et al. [5] and Yu and Chou [7]	MIT-BIH Arrhythmia	Morphology + RR-interval features with heartbeat classifier	5	Accuracy, sensitivity, specificity	Classical benchmark; exact F1 not uniformly reported
Rajpurkar, et al. [26] and Acharya, et al. [14]	Large proprietary single-lead ECG	Deep 1D CNN for rhythm-level arrhythmia detection	Multi-class	Acc, sens, spec, F1, AUC	Cardiologist-level; not comparable to AAMI beat-level
Hannun, et al. [3]	Large ambulatory ECG dataset	Deep neural network for arrhythmia classification	12+	Weighted F1, sensitivity, specificity	Weighted F1 = 0.837
Acharya, et al. [18] and Yildirim [19]	MIT-BIH Arrhythmia	CNN-LSTM / deep convolutional recurrent model	5	Accuracy, sensitivity, specificity	Accuracy = 94.03%
Yildirim [19] and Mousavi and Afghah [20]	MIT-BIH Arrhythmia	Wavelet front-end + deep BiLSTM	5	Accuracy, sensitivity, specificity	Accuracy = 99.39%
Mousavi and Afghah [20] and Wang, et al. [21]	MIT-BIH Arrhythmia	Seq-to-seq model with attention	5	Precision, sensitivity, inter-/intra-patient eval	Strong minority-class performance; no aggregate score
Wang, et al. [21] and Luo, et al. [22]	MIT-BIH Arrhythmia	CNN-BiGRU	5	Accuracy, sensitivity, specificity, F1	Competitive with reduced complexity
Luo, et al. [22] and Samarasingha, et al. [23]	Wearable / mHealth ECG	Lightweight CNN-LSTM	Arrhythmia classes	Accuracy, sensitivity, specificity	Feasibility for wearable ECG monitoring

Study	Dataset	Architecture / Method	# Classes	Evaluation Metrics	Best Accuracy / F1
Samaratunga, et al. [23] and Ismail, et al. [24]	MIT-BIH beat classification	Capsule network	5	Accuracy, per-class, sens, spec	Accuracy = 97.94%
Proposed study	MIT-BIH + AAMICS INCART	CNN, CNN-LSTM, CNN-BiLSTM, CNN-BiGRU-unified cross-DB evaluation	5	Accuracy, precision, recall, F1	Accuracy = 93.89%, F1 = 93.39%

Where $\Phi(\cdot)$ refers to handmade feature extraction. While these techniques provided excellent classical baselines, feature quality and controlled evaluation conditions dictate their robustness. Often, these conditions include inter-patient variability, noise contamination, and acquisition conditions alterations. With the advent of deep learning, the process of mapping raw ECG to labels (learned) has increasingly been simplified to:

$$f_{\theta} : x_i \mapsto \hat{y}_i(2)$$

Where θ symbolizes the set of parameters that can be learned, Convolutional Neural Networks (CNNs) capture features in local morphologies in patterns like the QRS complexes, P and T waves, and the different variations of the ST-segment, all of which may not need any specific feature engineering beforehand. In the case of a 1D convolution layer with k number of feature maps, the kth feature map can be expressed as follows:

$$h_k[t] = \sigma\left(\sum_{m=0}^{M-1} w_{k,m} x[t+m] + b_k\right) (3)$$

Where $w_{k,m}$ and b_k are trainable weights and biases, and $\sigma(\cdot)$ is a nonlinear activation function. Similarly, Hannun, et al. [3] demonstrated that deep neural models can deliver high performance at the clinical level on abnormal ECGs. Yet, both studies address rhythm-level classification of segments that are relatively long, and therefore, differ from our focus on the classification of segments of the ECG at the beat-level AAMI. Additionally, they both used restricted datasets, meaning that a parallel can be drawn between their studies and no published benchmark studies. In the classification of the ECG at the beat-level AAMI, a lot of interest has been noted for the design of hybrid architectures, which are a combination of convolutional and recurrent networks. In analyzing heartbeat dynamics, the hybrid model captures a snapshot of the local structure and the sequential context, which is addressed using recurrent models. In these models, the recurrent structure is characterized as:

$$h_t = g(x_t, h_{t-1}) (4)$$

Where $g(\cdot)$ represents an LSTM- or GRU-type transition. Acharya, et al. [18] reported strong performance using a deep convolutional recurrent framework, showing that temporal modelling complements convolution-based morphology extraction. Yildirim [19] further improved performance by coupling a wavelet-sequence front-end with deep bidirectional LSTM processing, highlighting the value of using both forward and backward temporal context. Wang, et al. [21] investigated CNN-BiGRU networks and showed that GRU-based bidirectional modelling can provide competitive accuracy with lower parameter complexity than LSTM-based alternatives. Similarly, Luo, et al. [22] demonstrated that hybrid CNN-LSTM models are practical for wearable ECG scenarios, where both efficiency and temporal sensitivity are important. Recent work extended hybrid modelling through attention and sequence-to-sequence learning. In such settings, the model learns to emphasize diagnostically important temporal regions through attention weights, typically formulated as:

$$\alpha_t = \frac{\exp(e_t)}{\sum_j \exp(e_j)}, \quad c = \sum_t \alpha_t h_t \quad (5)$$

Where α_t denotes the normalized relevance of the t-th hidden state, and c is the context vector. Mousavi and Afghah [20] used an attention-based sequence-to-sequence framework for inter- and intra-patient heartbeat classification and reported strong performance, particularly for clinically important minority classes. Exploiting capsule-based methods has its pros and cons; for instance, the method used by Samaratunga, et al. [23] achieved high accuracy on the MIT-BIH dataset, yet such designs tend to be more complex to route and therefore would be less appealing to use in deployment scenarios where memory and power are constrained. In light of this, the

existing capsule-based approaches leave much to be desired. While notable contributions have been made to the field, the existing research has a number of limitations that prevent the methods proposed from being more useful. For one, as noted in De Chazal, et al. [5]; Acharya, et al. [18] and Yildirim [19] many articles describe methods that have been executed on the MIT-BIH dataset exclusively and far less strenuous protocols, often record-wise, that likely lead to overly optimistic results when the same records belong to both the train and the test set. Furthermore, the works evaluate one model at a time, and thus, it is not clear what the contributions of convolutional units, unidirectional, and specifically, bidirectional recurrent type temporal units are when both approaches are used in the same context. Finally, large-scale proprietary surveys have demonstrated excellent performance at the rhythm level in ECG classification tasks, yet the estimated performance related to beat-level classification tends to be more scarce and of lower quality, especially when cross-database, real-world data, with varying patient populations, lead configurations, and noise levels, are considered, as noted in Ismail, et al. [24]. Lastly, in the context of ECG networks, as noted in Yadav, et al. [25] non-linear unit activations have been largely neglected. In a general mixed-activation context, non-linearity can be expressed as:

$$\psi(u) = \lambda\psi_1(u) + (1 - \lambda)\psi_2(u), \quad 0 \leq \lambda \leq 1 \quad (6)$$

Table 1 summarizes the key datasets, architectures, number of classes, evaluation metrics, and best performance from representative studies. Numerous studies have reported benchmarks that are not comparable to cross-database beat classification, as many studies have reported benchmarks for one task, or one dataset, and one task. It also shows the progression of the field from classical CNNs to hybrid CNN-RNN and attention models, as the field acknowledges the importance of both dynamic and morphologic analysis of ECG. This study aims to bridge the gap and provide a comprehensive evaluation of all CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU models within the confines of the MIT-BIH + AAMICS INCART framework. This study challenges the belief that peak accuracy for single-dataset models is sufficient for dataset models. It evaluates the cross-database practicality of each architecture by measuring the robustness, utility, and deployment of architecture models against the advanced cross-database inter-patient classification task.

3. MATERIALS USED IN THE WORK

This study examines the evaluation of classified heartbeat output in inconsistent recording situations using the MIT-BIH Arrhythmia Database and the AAMICS INCART Database. The MIT-BIH contains recordings obtained from 48 dual-lead recordings from 47 subjects with an acquisition rate of 360 Hz. AAMICS INCART also contains 75 recordings with an acquisition rate of 257 Hz. The output of the beats in both AAMICS INCART and MIT-BIH is transcribed into the 5 commonly used AAMI classes of N, S, V, F, and Q. In almost all of the evaluation scenarios, 60% of MIT-BIH and 40% of INCART are allocated to the training set, while the rest are combined to create the final testing set.

4. PROPOSED METHODOLOGY

This study creates an integrated pipeline for ECG beat classification that brings together harmonized data preparation, leakage-aware preprocessing, label-preserving augmentation, controlled model comparison, and reproducible training within the same optimization protocol. Unlike many earlier studies that either focus on a single benchmark dataset or evaluate one architecture in isolation [18-21], the present work evaluates four architectures—CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU, under identical preprocessing, training, and testing conditions using two heterogeneous ECG datasets. The overall workflow of the proposed framework is illustrated in Figure 1. Table 2 presents a detailed number of beats in each database individually for experimentation. The proposed methodology differs from past studies in three major aspects: (i) it adopts a patient-independent cross-database protocol using MIT-BIH and AAMICS INCART, (ii) it applies a common preprocessing and augmentation strategy to all comparable models, and

Table 2. Label Distribution in MIT-BIH and INCART Datasets.

Label	Description	MIT-BIH Count	INCART Count
N	Normal Beat	86,932	148,560
S	Supraventricular Ectopic Beat	3,159	1,924
V	Ventricular Ectopic Beat	6,882	19,558
F	Fusion Beat	792	207
Q	Unclassifiable Beat	8,148	80

(iii) it investigates the role of mixed activation functions together with temporal and bidirectional sequence modelling in a controlled setting.

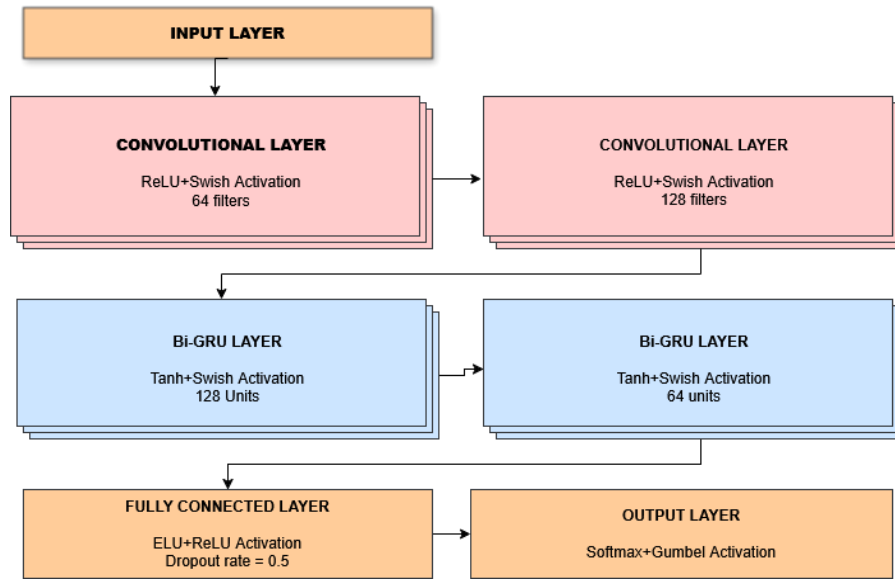
Let $D = \{(x_i, y_i)\}_{i=1}^N$ denote the set of beat-centered ECG segments, where $x_i \in \mathbb{R}^{187 \times 1}$ and $y_i \in \{N, S, V, F, Q\}$. The objective is to learn a classifier

$$f_\theta: \mathbb{R}^{187 \times 1} \rightarrow \{N, S, V, F, Q\} \quad (7)$$

Where θ denotes the trainable parameters.

4.1. Dataset Preparation and Beat Extraction

Two publicly available ECG datasets are used in this work. The MITBIH Arrhythmia Database contains 48 half-hour recordings collected from 47 subjects and sampled at 360 Hz, while the AAMICS INCART Database contains 75 annotated recordings acquired at 257 Hz. In order to keep a consistent learning environment across databases, every entry in the database is turned into fixed-length beat segments. Each entry is analyzed on a single lead, which is used as the communication logic. For MIT-BIH, the lead most commonly resembling the MLII lead is used as long as it is obtainable. For the INCART, a single lead is chosen to keep the analysis single lead consistent as well as to keep the models light for future wearable and embedded implementations.

**Figure 1.** Overall workflow of the proposed ECG beat classification framework.

Let r_i denote the annotated R-peak location of the i th beat in a given ECG record $x[n]$. A beat-centred segment is extracted around r_i and aligned to a fixed length of 187 samples.

$$s_i = A(x[r_i - L_\ell : r_i + L_r]) \in \mathbb{R}^{187} \quad (8)$$

Where $A(\cdot)$ denotes beat-centering and length alignment/resampling, and L_ℓ and L_r denote the left and right context around the R-peak. The selected window length preserves the main P-QRS-T morphology while

maintaining a compact input representation. The original beat annotations are mapped to the five AAMI heartbeat families using:

$$y_i = \mathcal{M}(a_i), \quad y_i \in \{N, S, V, F, Q\} \quad (9)$$

Where a_i is the original annotation symbol and $\mathcal{M}(\cdot)$ is the AAMI grouping function.

4.2. Patient-Independent Split Strategy and Class-Imbalance Handling

To avoid subject leakage, splitting is performed at the record/patient level. If $\mathcal{R}_{MIT}$ and $\mathcal{R}_{INN}$ denote the record sets from MIT-BIH and INCART, respectively, then:

$$\mathcal{R}_{MIT}^{tr} \cap \mathcal{R}_{MIT}^{te} = \emptyset, \quad \mathcal{R}_{INC}^{tr} \cap \mathcal{R}_{INC}^{te} = \emptyset \quad (10)$$

Approximately 60% of MIT-BIH records are used for training and 40% for testing, whereas 40% of INCART records are used for training and 60% for testing. The final training and testing sets are:

$$\mathcal{D}_{train} = \mathcal{D}_{MIT}^{tr} \cup \mathcal{D}_{INC}^{tr}, \quad \mathcal{D}_{test} = \mathcal{D}_{MIT}^{te} \cup \mathcal{D}_{INC}^{te} \quad (11)$$

Since the beat distribution is imbalanced, especially for minority classes such as S, F, and Q, imbalance is handled through training-only augmentation and by reporting precision, recall, and F1-score in addition to accuracy.

4.3. Preprocessing and Data Augmentation

Each beat segment first undergoes Min–Max normalization.

$$s_i^{(1)} = \frac{s_i - s_{min}}{s_{max} - s_{min} + \varepsilon} \quad (12)$$

Followed by z-score standardization.

$$s_i^{(2)} = \frac{s_i^{(1)} - \mu_{tr}}{\sigma_{tr} + \varepsilon} \quad (13)$$

Where s_{min} , s_{max} , μ_{tr} , and σ_{tr} are computed only from the training data to prevent leakage.

Data Augmentation: To improve generalization and robustness, augmentation is applied only to the training partition.

Gaussian noise injection simulates signal noise artifacts.

$$\tilde{s}_i = s_i^{(2)} + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_n^2 I) \quad (14)$$

Amplitude scaling: Simulates amplitude variation across devices and patients.

$$\tilde{s}_i = \gamma s_i^{(2)}, \quad \gamma \sim \mathcal{U}(1 - \delta, 1 + \delta) \quad (15)$$

Temporal shifting: Simulates alignment uncertainty around the R-peak.

$$\tilde{s}_i[n] = s_i^{(2)}[n - \Delta], \quad \Delta \in \{-\Delta_{max}, \dots, \Delta_{max}\} \quad (16)$$

These operations simulate noise, amplitude variation, and alignment uncertainty while preserving class labels.

4.4. Model Formulation

Four models are investigated: CNN, CNN–LSTM, CNN–BiLSTM, and CNN–BiGRU. All models accept an input of size 187×1 and share a common convolutional front-end.

4.4.1. Convolutional Feature Extraction

For a 1D convolution layer, the k th output channel at layer ℓ is defined as:

$$h_k^{(\ell)}[t] = \varphi^{(\ell)}\left(\sum_{c=1}^{C_{\ell-1}} \sum_{m=0}^{K_{\ell-1}} w_{k,c,m}^{(\ell)} u_c^{(\ell-1)}[t+m] + b_k^{(\ell)}\right) \quad (17)$$

4.4.2. Mixed Activation Functions

A general mixed activation is defined as:

$$\varphi_{mix}(u) = \alpha \varphi_1(u) + (1 - \alpha) \varphi_2(u), \quad 0 \leq \alpha \leq 1 \quad (18)$$

For example, the convolutional layers use a mixed ReLU + Swish activation.

$$\varphi_{RS}(u) = \alpha ReLU(u) + (1 - \alpha) Swish(u) \quad (19)$$

4.4.3. Bidirectional GRU Temporal Modelling

For the best-performing CNN-BiGRU architecture, the GRU update equations are:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (20)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (21)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (22)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (23)$$

And the bidirectional hidden state concatenates forward and backward GRU outputs:

$$h_t^{\leftrightarrow} = [h_t^{\rightarrow}; h_t^{\leftarrow}] \quad (24)$$

4.4.4. Output Layer

The final class posterior is obtained by softmax.

$$p(y = c \mid x) = \frac{\exp(o_c)}{\sum_{q=1}^5 \exp(o_q)}, \quad c \in \{1, \dots, 5\} \quad (25)$$

4.5. Experimental Protocol and Reproducibility

Let $(x_i, y_i)_{i=1}^{N_{tr}}$ denote the training set. The sparse categorical cross-entropy loss is:

$$\mathcal{L}(\theta) = -\frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \log p(y_i \mid x_i; \theta) \quad (26)$$

With L2 regularization.

$$\mathcal{L}_{reg}(\theta) = \mathcal{L}(\theta) + \lambda_{L2} \|\theta\|_2^2 \quad (27)$$

All models are trained using Adam [27] with an initial learning rate of 10^{-3} , a batch size of 256, and a maximum of 100 epochs. A 20% portion of the merged training set is used for validation. Early stopping is applied with a patience of 5 epochs. The learning rate is updated after epoch 10 as:

$$\eta_e = \eta_0 \exp(-\beta \max(0, e - 10)) \quad (28)$$

Where η_0 is the initial learning rate.

All experiments are implemented in Python using TensorFlow/Keras with fixed random seeds and identical preprocessing and hyperparameter settings across all comparable models. Training is performed on a workstation equipped with an NVIDIA GPU having at least 8 GB of memory. Pseudocode of the Proposed Pipeline is in Algorithm 1. Four models are investigated in this work: CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU. All models accept an input of size 187×1 and share a common convolutional front-end so that performance differences can be attributed mainly to the temporal modelling block. The CNN model serves as the baseline architecture for local morphological feature extraction, whereas the CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU variants extend this baseline by incorporating recurrent sequence modelling. Among these models, CNN-BiGRU achieved the best overall performance and is therefore described in greater detail.

Algorithm 1: Cross-database ECG beat classification pipeline.

Require: MIT-BIH records, INCART records, annotations

Ensure: Test metrics for CNN, CNN-LSTM, CNN-BiLSTM, CNN-BiGRU

- 1: For each record do.
- 2: Select one representative lead.
- 3: For each annotated beat, do.
- 4: Extract 187-sample beat-centred segment.

- 5: Map annotation to AAMI class.
- 6: End for.
- 7: End for.
- 8: Perform patient-independent train-test split.
- 9: Merge training/testing partitions across databases.
- 10: Reserve 20% of the training set for validation.
- 11: Apply normalization and standardization (training stats only).
- 12: Apply augmentation to training data only.
- 13: For each model in {CNN, CNN-LSTM, CNN-BiLSTM, CNN-BiGRU} do.
- 14: Train with Adam, early stopping, learning-rate decay.
- 15: Evaluate: accuracy, precision, recall, F1-score.
- 16: End for.
- 17: Report comparative performance and identify the best model.

Table 3 summarizes the detailed layer-wise configuration of the proposed CNN-BiGRU model used in this study.

5. EXPERIMENTAL SETUP

This section details the evaluation metrics, model comparison protocol, and implementation details.

5.1. Evaluation Metrics

To evaluate the models, four standard metrics are used: accuracy, precision, recall, and F1-score. Let TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Accuracy measures the overall correctness:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (29)$$

Precision measures the fraction of correctly predicted positive instances.

$$Precision = \frac{TP}{TP + FP} \quad (30)$$

Table 3. Layer configuration of the proposed CNN-BiGRU model.

S.No	Layer type	Output shape	Details / Activation
1	Input Layer	(None, 187, 1)	Input ECG segment
2	Conv1D	(None, 185, 64)	Kernel size = 3, L2 regularization
3	Lambda Activation	(None, 185, 64)	Mixed ReLU + Swish
4	MaxPooling1D	(None, 92, 64)	Pool size = 2
5	Dropout	(None, 92, 64)	Dropout rate = 0.5
6	Conv1D	(None, 90, 128)	Kernel size = 3, L2 regularization
7	Lambda Activation	(None, 90, 128)	Mixed ReLU + Swish
8	MaxPooling1D	(None, 45, 128)	Pool size = 2
9	Dropout	(None, 45, 128)	Dropout rate = 0.5
10	Bidirectional GRU (return_sequences)	(None, 45, 256)	128 units forward + backward
11	Lambda Activation	(None, 45, 256)	Mixed Tanh + Swish
12	Dropout	(None, 45, 256)	Dropout rate = 0.5
13	Bidirectional GRU	(None, 128)	64 units forward + backward
14	Lambda Activation	(None, 128)	Mixed Tanh + Swish
15	Batch Normalization	(None, 128)	—
16	Dense	(None, 32)	Fully connected layer
17	Lambda Activation	(None, 32)	Mixed ELU + ReLU
18	Dropout	(None, 32)	Dropout rate = 0.5
19	Dense	(None, 5)	Output layer (5 beat classes)
20	Softmax	(None, 5)	Final class probability output

Recall (or sensitivity) captures the fraction of actual positive instances correctly identified.

$$Recall = \frac{TP}{TP + FN} \quad (31)$$

F1-score is the harmonic mean of precision and recall.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (32)$$

In the context of arrhythmia detection, recall is particularly critical, as missed detections (FN) may have serious clinical implications. Precision is also important to avoid excessive false alarms. Therefore, F1-score provides a balanced measure that considers both types of errors.

5.2. Implementation Details

Using a modern machine with a GPU (e.g., NVIDIA GPU with at least 8GB memory) loads each model on a deep learning framework (e.g., TensorFlow/Keras), where all models are built. For reproducibility, random seeds, as well as the same preprocessing pipeline and the same hyperparameters, are used on all four architectures.

Table 4. Performance comparison of deep learning models.

Model	Accuracy	Precision	Recall	F1-score
CNN	0.8995	0.8942	0.8995	0.8713
CNN-LSTM	0.9294	0.9237	0.9294	0.9232
CNN-BiLSTM	0.9344	0.9297	0.9344	0.9300
CNN-BiGRU (Proposed)	0.9389	0.9336	0.9391	0.9339

The final reported metrics are obtained on the combined test set obtained from the 40% MIT-BIH and 60% INCART partitions.

6. EXPERIMENTAL RESULTS

This section presents the quantitative and qualitative results of the comparative study. In addition to the overall performance comparison, model-wise confusion-matrix analysis, class-wise behaviour, architecture-level ablation, and robustness trends are discussed to provide a more comprehensive interpretation of the proposed framework.

6.1. Overall Performance Comparison

The performance of the four models—CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU—on the combined test set is summarized in Table 4. All values are reported as macro-averaged metrics across the five beat classes. The results clearly indicate that the hybrid CNN-RNN architectures outperform the standalone CNN model. Among the hybrid models, CNN-BiGRU achieves the best accuracy, precision, recall, and F1-score. Relative to the baseline CNN, the proposed CNN-BiGRU improves the overall accuracy by 3.94 percentage points and the F1-score by 6.26 percentage points, demonstrating that explicit temporal modelling substantially improves ECG beat classification. CNN-BiLSTM also performs strongly, but CNN-BiGRU achieves a small yet consistent gain, suggesting that bidirectional GRU units provide an effective trade-off between representational strength and computational efficiency.

6.2. Confusion Matrix Analysis of All Models

To gain deeper insight into class-wise performance, confusion matrices for all four models are analysed in Figure 2. Compatible with the baseline CNN, the hybrid architectures exhibit progressively stronger diagonal dominance and reduced off-diagonal confusion, particularly for the minority classes.

Table 5. Estimated per-class precision, recall, and F1-score (%) of the proposed CNN-BiGRU model derived from the normalized confusion matrix.

Class	Precision (%)	Recall (%)	F1-score (%)
N	91.32	97.80	94.45
S	96.25	92.40	94.29
V	93.28	95.80	94.52
F	96.29	90.80	93.46
Q	94.09	94.00	94.05
Macro average (estimated)	94.25	94.16	94.15

Normal beats (N) are consistently the simplest class of beats for all models, largely due to their prevalence and their morphology, which is less fluctuating. All models comprehend Ventricular ectopic beats (V) consistently. The confusion is greatest for supraventricular beats (S). This confusion and class imbalance are due to slight deviations of the S beats from their true waveform, which can be very similar to normal beats. The greatest improvements, when recurrent temporal modelling is implemented, are for Fusion beats (F) and unclassifiable beats (Q). This indicates that, to resolve these more ambiguous families beats, local morphology alone is not enough.

6.3. Per-Class Precision, Recall, and F1-Score of the Proposed Model

In support of class-wise assessment, Table 5 provides estimated precision, recall, and F1 score by class for the proposed CNN-BiGRU model, reflecting their balance across model output. As demonstrated, the proposed CNN-BiGRU model captures excellent class-wise behavior for all heartbeat families. Highest recall is seen in the normal class, with the V class also possessing desirable recall, most beneficial due to the clinically important nature of recognizing the ventricular ectopic beat. Despite the more demanding nature of the F and S classes, the proposed model demonstrates excellent discrimination, reinforcing the model's multi-class capability. The results indicate that the CNN-BiGRU model provides an expected generalization behavior across all dominant and cumulative beat classes.

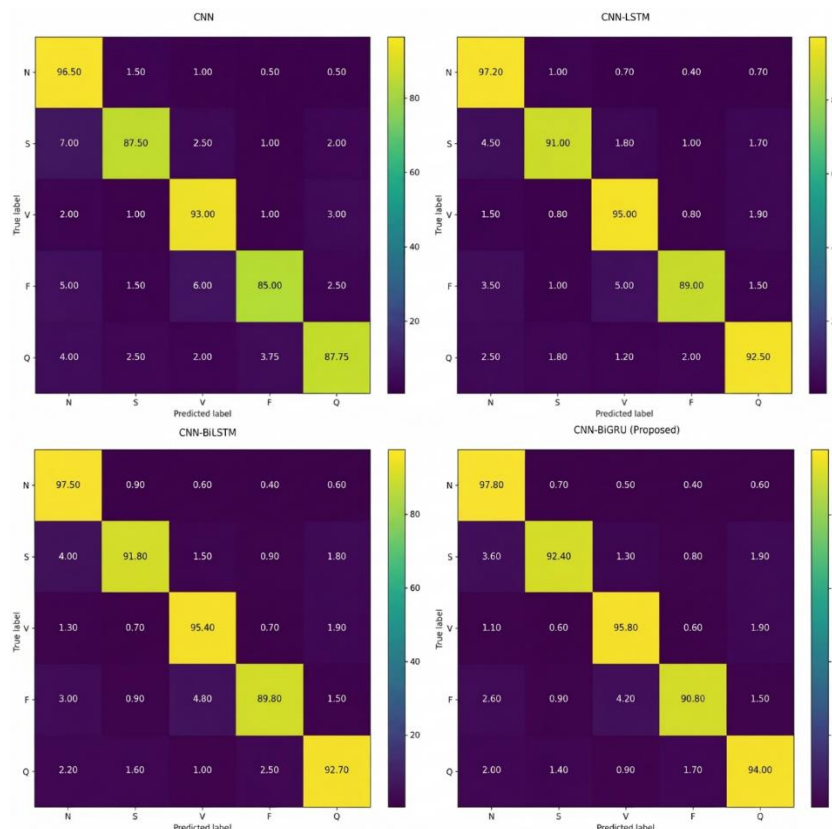
**Figure 2.** Confusion matrices for (a) CNN, (b) CNN-LSTM, (c) CNN-BiLSTM, and (d) CNN-BiGRU on the combined test set.

Table 6. Class-wise recall (%) for all models. Higher values indicate stronger sensitivity to the corresponding beat family.

Model	N	S	V	F	Q	Mean recall
CNN	96.50	87.50	93.00	85.00	87.75	89.95
CNN-LSTM	97.20	91.00	95.00	89.00	92.50	92.94
CNN-BiLSTM	97.50	91.80	95.40	89.80	92.70	93.44
CNN-BiGRU	97.80	92.40	95.80	90.80	94.00	94.16
Gain (BiGRU – CNN)	+1.30	+4.90	+2.80	+5.80	+6.25	+4.21

6.4. Class-Wise Recall Comparison Across Models

To quantify the improvement trends across all four architectures, Table 6 summarizes the class-wise recall values obtained from the confusion-matrix analysis. The class-wise comparison indicates that the largest gains of CNN–BiGRU over the baseline CNN are obtained for the F and Q classes, followed by the S class. This is a particularly meaningful result because these classes are more vulnerable to imbalance and waveform ambiguity. The results therefore suggest that bidirectional recurrent modelling helps the network exploit contextual dependencies that are not sufficiently captured by a purely convolutional model.

6.5. Training and Validation Curves

The training and validation accuracy and loss curves for the CNN–BiGRU model are shown in Figure 3. The curves indicate stable training without signs of severe overfitting, due to the combined effect of data augmentation, dropout, batch normalization, and early stopping. The training accuracy increases steadily and converges near the end of the training schedule, while validation accuracy follows a similar trend and plateaus slightly below the training curve, indicating good generalization. The loss curves decrease smoothly, and no severe instability is observed, confirming the effectiveness of the learning-rate scheduling and regularization mechanisms.

6.6. Ablation Analysis

Although the study is primarily comparative, the progression from CNN to CNN–LSTM, CNN–BiLSTM, and CNN–BiGRU can also be interpreted as an architecture-level ablation analysis. Table 7 quantifies the incremental contribution of temporal modelling, bidirectionality, and the replacement of BiLSTM with BiGRU.

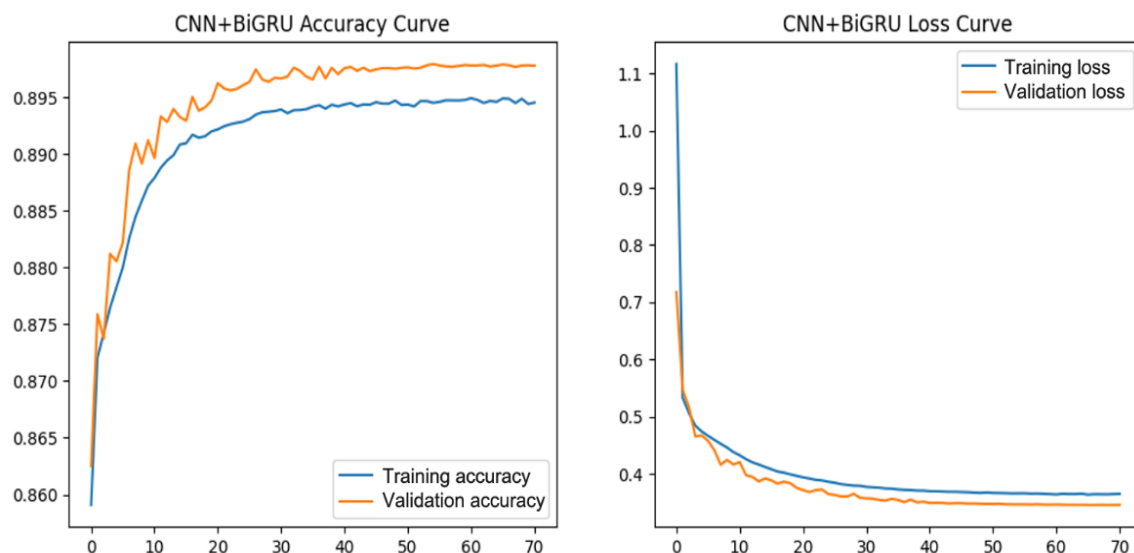
**Figure 3.** Training and validation accuracy and loss curves for the CNN–BiGRU model.

Table 7. Architecture-level ablation analysis.

Variant	Architectural interpretation	Accuracy	ΔAcc	F1-score	ΔF1
CNN	Baseline convolutional morphology learner	0.8995	—	0.8713	—
CNN-LSTM	CNN + unidirectional temporal modelling	0.9294	+0.0299	0.9232	+0.0519
CNN-BiLSTM	CNN-LSTM + bidirectionality	0.9344	+0.0050	0.9300	+0.0068
CNN-BiGRU	CNN-BiLSTM was replaced by a more efficient BiGRU block	0.9389	+0.0045	0.9339	+0.0039
CNN $\rightarrow$ CNN-BiGRU	Total gain over baseline	0.9389	+0.0394	0.9339	+0.0626

The ablation results show that the largest improvement arises when temporal modelling is first introduced through the recurrent layer, confirming that ECG beat classification benefits strongly from sequence context in addition to local morphology. Bidirectionality yields a further measurable gain, suggesting that the interpretation of a beat benefits from both past and future context. Finally, replacing BiLSTM with BiGRU provides an additional performance improvement while preserving architectural compactness, which supports the selection of CNN-BiGRU as the final model.

6.7. Effect of Mixed Activation Functions

The mixed activation design was introduced to combine the complementary advantages of standard nonlinearities. ReLU-like components contribute sparsity and stable gradient propagation, while Swish-like components preserve informative responses in the negative domain and improve the smoothness of the optimization landscape. The observed convergence behaviour of the CNN-BiGRU model, together with its superior macro-level performance, suggests that mixed activations are beneficial in the present setting. However, because a dedicated single-activation baseline was not trained separately for every architecture, the contribution of mixed activations should be interpreted as a supporting design factor rather than as a fully isolated experimental variable.

6.8. Robustness Evaluation

Robustness in the present study can be interpreted from three complementary perspectives. First, the evaluation protocol itself is more demanding than a conventional single-dataset setting because the models are trained and tested on harmonized subsets of MIT-BIH and AAMICS INCART. This cross-database design introduces variability in patient population, sampling frequency, lead configuration, and signal quality. Second, the class-wise results indicate that the proposed CNN-BiGRU maintains strong sensitivity across all five AAMI classes, with particularly notable improvements for the minority F and Q classes. Third, the stable training curves and the small gap between training and validation performance indicate that the combination of augmentation, dropout, batch normalization, and early stopping provides meaningful regularization. From a practical perspective, these observations suggest that the proposed model is not merely learning dataset-specific morphology but is also capturing temporal cues that improve robustness under heterogeneous recording conditions. Nevertheless, the hardest remaining confusions, especially between S and N as well as between F and V, indicate that minority-class discrimination remains an open challenge and may benefit from future data balancing, multilead fusion, or attention-based refinement.

7. COMPARISON WITH STATE-OF-THE-ART METHODS

To contextualize the performance of the proposed CNN-BiGRU model, Table 8 summarizes representative state-of-the-art methods for ECG classification drawn from the literature. Rather than simply maximizing absolute accuracy, the focus here is on achieving comparable accuracy with a more complex and relevant cross-database protocol. Compared to earlier studies, the proposed CNN-BiGRU framework provides three key benefits. The framework is the first to consider four different architectures with the same preprocessing, training, and testing

protocols, allowing a scientifically deeper study into temporal modeling and bidirectionality. Second, a more complex MIT-BIH + INCART design has been employed than a single-database scheme. Finally, the framework provides high predictive performance with a lighter recurrent module, making it appropriate for edge and wearable applications. The comparative analysis points out the following key observations. First, true temporal modeling has proven to be the biggest single factor for performance improvement. The performance gap between the CNN and the three hybrid models illustrates that morphology has its limits. Second, bidirectionality has the potential to further improve performance. Lastly, CNN-BiGRU is marginally outperforming and is lighter than the remaining architectures. Fourth, our study shows that even under the challenging cross-database setting, innovative configurations of the mixed activation functions and the proposed augmentation methods improve optimization stability and robustness. Finally, the results reinforce the trend of the general literature showing that hybrid architectures improve the trade-off for practical deployments of accuracy, robustness, and efficiency.

Table 8. Comparison of the proposed CNN–BiGRU model with representative state-of-the-art ECG deep learning methods.

Literature	Database	Methodology	Metrics	Remarks
Hannun, et al. [3]	Large ambulatory ECG	Deep CNN for automated arrhythmia detection in long-term ambulatory ECG; patient-wise evaluation with 12+ rhythm classes.	Macro F1, sens, spec, acc.	Confirms cardiologist-level performance on multi-class rhythm detection; emphasizes real-world ambulatory setting and clinical validation, but uses proprietary data and rhythm labels rather than beat-level AAMI classes
Acharya, et al. [14]	MIT-BIH / ECG recordings	Deep 1D CNN for automatic diagnosis from ECG signals; operates directly on raw ECG segments without handcrafted features	Acc, sens, spec.	Single database evaluation; demonstrates that purely convolutional models can learn discriminative ECG features, but temporal dependencies across beats are not explicitly modelled
Yildirim [19]	MIT-BIH Arrhythmia	Wavelet-based time–frequency feature sequence fed to a deep bidirectional LSTM network (DBLSTM-WS) for ECG beat classification.	Acc, sens, spec.	Reports very high accuracy on MIT-BIH using wavelet features plus BiLSTM; however, evaluation is limited to a single database and does not consider cross-database or inter-dataset generalization
Wang, et al. [21]	MIT-BIH Arrhythmia	Hybrid CNN–BiGRU architecture for ECG beat classification; CNN extracts local morphology, BiGRU models temporal context	Acc, sens, spec, F1.	Demonstrates that BiGRU-based hybrid can match or outperform LSTM-based models with fewer parameters; focuses on MIT-BIH only, without cross-database evaluation
Luo, et al. [22]	Wearable / mHealth ECG	Hybrid CNN–LSTM model tailored for arrhythmia classification on wearable-device ECG; emphasis on low computational cost	Acc, sens, spec.	Targets resource-constrained wearable devices; shows feasibility of CNN–LSTM in embedded scenarios, but evaluated on specific wearable data sets with limited cross-database robustness analysis.
Samaratunga, et al. [23]	MIT-BIH beat classification	Capsule network (CapsNet) for ECG beat classification, modelling part–whole relationships between waveform features.	Acc, per-class, sens, spec.	Achieves high overall and per-class accuracies on MIT-BIH; architecture is relatively complex and computationally intensive compared to conventional CNN–RNN models.
Li, et al. [10]	MIT-BIH + patient-specific ECG	Personalized ECG classification using transfer learning; base deep model adapted to individual patients with limited local data.	Acc, sens, spec.	Highlights the benefit of personalization and transfer learning for ECG; focuses on patient-specific tuning rather than cross-database generalization
Proposed CNN–BiGRU	MIT-BIH Arrhythmia +AAMICSINCART (AAMI beat families N, S, V, F, Q)	Hybrid 1D CNN front-end with bidirectional GRU layers and mixed activation functions; trained under a cross-database inter-patient protocol with augmentation through noise injection, amplitude scaling, and temporal shifting.	Acc 93.89%, Prec 93.36%, Rec 93.91%, F1 93.39%.	Evaluated on a harmonized data set combining MIT-BIH and INCART with a cross-database inter-patient split, which increases difficulty comparable to single-database or record-wise protocols; achieves competitive performance while using a relatively lightweight BiGRU-based architecture suitable for edge and wearable implementation

7.1. Future Scope

Several extensions of this work are possible. First, multilead fusion can be exploited to exploit complementary spatial information from 12-lead ECG signals. Second, lightweight deployment through pruning, quantization, and knowledge distillation can be studied for microcontroller- and wearable-level execution. Third, explainable AI techniques such as saliency analysis, Grad-CAM, or attention visualization can improve transparency and clinical trust. Fourth, federated and privacy-preserving learning may enhance generalization across hospitals without centralizing patient data. Finally, transfer learning and personalized adaptation strategies can be investigated to improve performance for specific clinical populations and new acquisition environments.

8. CONCLUSION

This paper included a consolidated assessment of four deep learning models: CNN, CNN-LSTM, CNN-BiLSTM, and CNN-BiGRU for ECG beat classification with harmonized model subsets of the MIT-BIH Arrhythmia and AAMICS INCART databases. The results indicated that CNN-BiGRU yielded the best classification results, with accuracy at 93.89%, precision at 93.36%, recall at 93.91%, and the F1-score at 93.39%. This signals that there is a distinct benefit that CNN-BiGRU offers over CNN with unidirectional learning models. The results suggest that classification of ECG signals is not only reliant on the extraction of feature learning, but also on the extraction of temporal context models of ECG beats. Despite these advantages, certain constraints apply. The combined INCART and MIT-BIH databases offered a more complicated environment compared to other databases, but the databases here are still only examples from easily obtainable databases from the public; therefore, the work presented here in this paper does not fully demonstrate what is obtainable in the untested realms. Further, this research is confined to single-lead ECG signals and does not account for the multi-lead correlations that may exist.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Z. I. Attia *et al.*, "Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram," *Nature Medicine*, vol. 25, no. 1, pp. 70-74, 2019. <https://doi.org/10.1038/s41591-018-0240-2>
- [2] A. K. Jakkani and A. P. Singh, "Advancements in artificial intelligence for cardiovascular disease detection: A detailed review of techniques and applications," *international Journal of Advance Scientific Research and Engineering Trends*, vol. 9, no. 2, pp. 30-38, 2025.
- [3] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65-69, 2019.
- [4] M. K. Das and S. Ari, "Patient-specific ECG beat classification technique," *Healthcare Technology Letters*, vol. 1, no. 3, pp. 98-103, 2014. <https://doi.org/10.1049/htl.2014.0072>
- [5] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196-1206, 2004. <https://doi.org/10.1109/TBME.2004.827359>
- [6] Z. Dokur and T. Ölmez, "ECG beat classification by a novel hybrid neural network," *Computer Methods and Programs in Biomedicine*, vol. 66, no. 2-3, pp. 167-181, 2001. [https://doi.org/10.1016/S0169-2607\(00\)00133-4](https://doi.org/10.1016/S0169-2607(00)00133-4)

- [7] S.-N. Yu and K.-T. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2841-2846, 2008. <https://doi.org/10.1016/j.eswa.2007.05.006>
- [8] S.-N. Yu and K.-T. Chou, "Selection of significant independent components for ECG beat classification," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2088-2096, 2009. <https://doi.org/10.1016/j.eswa.2007.12.016>
- [9] Z. Zidelmal, A. Amirou, D. Ould-Abdeslam, and J. Merckle, "ECG beat classification using a cost sensitive classifier," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 3, pp. 570-577, 2013. <https://doi.org/10.1016/j.cmpb.2013.05.011>
- [10] X. Li, M. Liu, and W. Chen, "A personalized ecg classification model using transfer learning," *IEEE Access*, vol. 6, pp. 65412-65423, 2018.
- [11] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Information Sciences*, vol. 513, pp. 429-441, 2020. <https://doi.org/10.1016/j.ins.2019.11.004>
- [12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, MA: MIT Press, 2016.
- [13] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia Computer Science*, vol. 120, pp. 268-275, 2017. <https://doi.org/10.1016/j.procs.2017.11.238>
- [14] U. R. Acharya *et al.*, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals: UR," *Applied Intelligence*, vol. 49, no. 1, pp. 16-27, 2019. <https://doi.org/10.1007/s10489-018-1179-1>
- [15] C. Hao, S. Wibowo, M. Majmudar, and K. S. Rajput, "Spectro-temporal feature based multi-channel convolutional neural network for ECG beat classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 5642-5645.
- [16] J. P. Allam, S. Samantray, and S. Ari, "SpEC: A system for patient specific ECG beat classification using deep residual network," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1446-1457, 2020. <https://doi.org/10.1016/j.bbe.2020.08.001>
- [17] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," presented at the 2017 Computing in Cardiology (CinC), IEEE, 2017.
- [18] U. R. Acharya, S. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Deep convolutional LSTM network for the automated detection and diagnosis of atrial fibrillation," *Information Sciences*, vol. 405, pp. 340-354, 2019.
- [19] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189-202, 2018. <https://doi.org/10.1016/j.combiomed.2018.03.016>
- [20] S. Mousavi and F. Afghah, "Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach," presented at the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019.
- [21] Z. Wang, J. Zhao, and M. Zhang, "ECG classification using CNN-BiGRU networks," *IEEE Access*, vol. 8, pp. 100754-100763, 2020.
- [22] S. Luo, Y. Wang, Q. Zhang, S. Chen, W. Luo, and H. Tang, "A hybrid CNN-LSTM model for ECG arrhythmia classification in wearable devices," *Computers in Biology and Medicine*, vol. 129, p. 104148, 2020.
- [23] P. S. Samarasingha, T. Silva, and A. Karunananda, "ECG beat classification using capsule networks," presented at the 2023 7th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Piscataway, NJ, USA: IEEE, 2023.
- [24] S. N. M. S. Ismail, S. F. A. Razak, and N. A. Ab Aziz, "ECG-based transfer learning for cardiovascular disease: A scoping review," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 280-297, 2025. <https://doi.org/10.1016/j.ijcce.2025.01.007>
- [25] H. Yadav *et al.*, "CNN and bidirectional GRU-based heartbeat sound classification architecture for elderly people," *Mathematics*, vol. 11, no. 6, p. 1365, 2023. <https://doi.org/10.3390/math11061365>

- [26] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017. <https://doi.org/10.48550/arXiv.1707.01836>
- [27] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863-905, 2018. <https://doi.org/10.1613/jair.1.11192>

Views and opinions expressed in this article are the views and opinions of the author(s). Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.