Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 1, February 2026, pp. 712~722
ISSN: 2302-9285, DOI: 10.11591/eei.v15i.10893

Comparative analysis of unidirectional and bidirectional RNNs for ECG arrhythmia detection using augmented MIT-BIH data

Sabura Banu Urundai Meeran[1], Nafeena Abdul Munaf[2]
[1]Department of Electrical and Electronics Engineering, Saveetha Engineering College, Chennai, India
[2]Faculty of Engineering and Technology, Villa College, Malé, Maldives

Article Info

Article history:

Received Jun 21, 2025
Revised Oct 4, 2025
Accepted Dec 6, 2025

Keywords:

Arrhythmia detection
Bidirectional recurrent neural network
Data augmentation
Electrocardiogram classification
Gated recurrent unit
MIT-BIH arrhythmia database

ABSTRACT

Accurate classification of electrocardiogram (ECG) signals is essential for early arrhythmia detection. This study compares the performance of unidirectional and bidirectional recurrent neural networks (RNN), specifically gated recurrent unit (GRU)-based architectures, for classifying ECG beats as normal or arrhythmic. ECG data were sourced from the MIT-BIH Arrhythmia Database using the WFDB toolkit. Each beat was segmented into a 128-sample window centered on the R-peak and labeled into two classes. To address severe class imbalance (6,279 normal vs. 43 arrhythmic beats), data augmentation techniques—jittering and scaling—were applied, resulting in a balanced dataset. Both models were trained under identical conditions, with evaluation based on accuracy, precision, recall, F1-score, and other statistical metrics. The unidirectional RNN achieved poor recall (9.0%) despite high precision, yielding an overall accuracy of 54.0%. In contrast, the bidirectional RNN significantly outperformed, achieving 98.17% accuracy, 98.39% precision, 97.92% recall, and a 98.16% F1-score. The results demonstrate that bidirectional temporal modeling provides substantial improvements in ECG classification, especially for detecting minority class arrhythmias. This study highlights the importance of both data augmentation and model architecture in developing effective deep learning solutions for real-time ECG analysis and clinical diagnostics.

Corresponding Author:

Sabura Banu Urundai Meeran
Department of Electrical and Electronics Engineering, Saveetha Engineering College
Thandalam, Chennai 602105, Tamil Nadu, India
Email: saburabanuu@saveetha.ac.in

1. INTRODUCTION

Cardiovascular diseases (CVDs) remain the foremost cause of global mortality, accounting for approximately 17.9 million deaths each year, according to the World Health Organization [1]. Among CVDs, cardiac arrhythmias constitute a major contributor to sudden cardiac arrest, long-term morbidity, and reduced quality of life. Early and reliable detection of arrhythmias is therefore crucial for effective clinical intervention and patient management. The electrocardiogram (ECG) is the primary diagnostic tool for identifying cardiac rhythm abnormalities; however, conventional manual ECG interpretation is labor-intensive, time-consuming, and prone to both intra- and inter-observer variability, even among experienced cardiologists [2], [3]. These limitations have motivated the development of automated ECG analysis systems aimed at improving diagnostic accuracy, consistency, and efficiency. A comprehensive systematic review of deep learning techniques applied to ECG-based arrhythmia classification was done [4]. The study analyzes various neural network architectures, datasets, and performance metrics, highlighting the superiority of deep learning models over traditional methods in detecting complex cardiac abnormalities. The authors also discuss existing challenges such as data imbalance, model interpretability, and the need for clinically validated datasets for real-world deployment.

Early automated arrhythmia detection approaches predominantly relied on handcrafted feature extraction combined with classical machine learning classifiers such as support vector machines (SVMs) [5], k-nearest neighbors (k-NN) [6], and decision tree-based models [7]. Although these techniques demonstrated reasonable performance in controlled settings, their effectiveness is constrained by the quality of manually engineered features and their limited ability to capture complex morphological variations and nonlinear temporal dynamics present in ECG signals [8]. Consequently, their generalization to diverse patient populations and real-world clinical data remains challenging.

Recent advances in deep learning have significantly transformed ECG signal analysis by enabling end-to-end learning directly from raw or minimally processed signals. Ferretti *et al.* [9] demonstrated the effectiveness of a 1D convolutional neural networks (CNN) model for arrhythmia classification, achieving high accuracy by learning spatial features from ECG signals without manual feature engineering. CNNs, in particular, have been widely adopted due to their strong capability to automatically learn discriminative spatial and morphological features from ECG waveforms [10], [11]. Ribeiro *et al.* [12] demonstrated cardiologist-level performance using deep neural networks for automatic ECG diagnosis, highlighting the clinical potential of CNN-based approaches. Nevertheless, CNNs primarily focus on local receptive fields and are inherently limited in modeling long-range temporal dependencies, which are essential for capturing rhythm-level patterns and beat-to-beat variability in ECG recordings [13]. To address temporal modeling limitations, recurrent neural networks (RNNs) have been extensively explored for ECG classification tasks. Architectures such as long short-term memory (LSTM) networks and gated recurrent units (GRUs) are specifically designed to learn sequential dependencies in time-series data, making them well-suited for ECG analysis [14]–[16]. Recent studies have reported improved classification performance using LSTM- and GRU-based models, including micro-class and multi-task learning strategies that enhance sensitivity to subtle arrhythmic patterns [17]–[19]. However, conventional unidirectional RNNs process signals only in the forward temporal direction, which may limit contextual understanding of cardiac cycles.

Bidirectional recurrent neural networks (Bi-RNNs) overcome this limitation by processing sequences in both forward and backward directions, thereby leveraging past and future contextual information simultaneously [20]. Such architectures have achieved remarkable success in speech recognition and sequence modeling tasks [21]. To address this limitation, RNNs, particularly long short-term memory (LSTM) architectures, have been explored for modeling ECG sequences. A bidirectional LSTM (Bi-LSTM) model that processes ECG signals in both forward and backward directions, allowing the network to capture richer temporal context and improve classification performance was proposed [22]. Building on this idea, a hybrid CNN–BiLSTM framework that combines spatial feature extraction with bidirectional temporal modeling, achieving enhanced diagnostic accuracy for arrhythmia detection was introduced [23]. Despite these improvements, class imbalance remains a major challenge in ECG datasets, where abnormal beats are often underrepresented. The issue using a micro-class approach that improves sensitivity to rare arrhythmia types was addressed [24]. Their results highlight the importance of designing models and training strategies that can effectively detect minority-class events, which are clinically critical for early diagnosis and intervention. This imbalance often biases learning algorithms toward majority classes, reducing sensitivity to clinically critical but infrequent arrhythmias. To mitigate this issue, various data augmentation and resampling techniques have been proposed, including the synthetic minority over-sampling technique (SMOTE) [25], generative adversarial networks (GANs) [26], and domain-specific signal transformations such as jittering, scaling, and amplitude modulation [27], [28]. While GAN-based approaches can generate highly realistic synthetic ECG signals, simpler augmentation methods remain computationally efficient and effective for enhancing model robustness, especially when combined with hybrid deep learning architectures [29].

Despite advances in deep learning for ECG classification, three critical gaps remain:
- Limited comparative studies on unidirectional vs. Bi-RNN under class-balanced conditions achieved through simple yet effective augmentation strategies.
- Under exploration of augmentation techniques specifically optimized for temporal deep learning models in ECG analysis.
- Lack of comprehensive evaluation metrics—such as receiver operating characteristic – area under the curve (ROC-AUC), Matthews correlation coefficient (MCC), and Cohen's Kappa—beyond accuracy, particularly for minority-class performance.

This study addresses these gaps by conducting a systematic comparison of GRU-based unidirectional and bidirectional architectures on an augmented MIT-BIH dataset. Jittering and scaling are applied to balance the dataset, and models are evaluated using an extensive set of performance metrics, including accuracy, precision, recall, F1-score, ROC-AUC, MCC, and Cohen's Kappa. The contributions of this work are as follows:

– Demonstrating the superiority of bidirectional temporal modeling for arrhythmia detection under balanced data conditions.
– Validating computationally efficient augmentation strategies for ECG classification.
– Providing a comprehensive benchmark against recent state-of-the-art methods.

The rest of the paper is organized as follows: section 2 details the materials and method, section 3 presents the results and discussion, and section 4 concludes with implications and future research directions.

2. MATERIALS AND METHOD

The methodology framework was designed to systematically evaluate the efficacy of unidirectional and bidirectional GRU models for ECG arrhythmia classification, leveraging the MIT-BIH Arrhythmia Database as the primary data source. ECG beats were extracted using the WFDB toolkit, with each heartbeat segmented into a 128-sample window centered on the R-peak, following established preprocessing protocols [24]. To address the severe class imbalance (6,279 normal vs. 43 arrhythmic beats), two computationally efficient augmentation techniques—jittering (addition of Gaussian noise) and scaling (amplitude modulation)—were applied to the minority class, balancing the dataset to 6,279 samples per class, as validated in recent works ([27], [28]). Both GRU architectures were constructed with identical hyperparameters (two layers, 64 hidden units, and Adam optimizer) and trained on 80% of the augmented data, while 20% was reserved for testing. Performance metrics, including accuracy, precision, recall, and F1-score, were computed to ensure comprehensive evaluation, with emphasis on minority-class detection. The bidirectional GRU processed sequences in forward and backward directions to capture contextual dependencies, contrasting with the unidirectional model's restricted temporal scope ([28], [29]). This experimental design ensures a fair comparison, isolating the impact of bidirectional temporal modeling while maintaining uniformity in data handling, training protocols, and evaluation criteria.

2.1. Dataset and preprocessing

The dataset used in this study is derived from the MIT-BIH Arrhythmia Database, a widely utilized and benchmarked resource for ECG signal analysis. Access to the dataset was facilitated through the WFDB Python toolkit, which enables efficient reading and manipulation of ECG recordings and annotations.

This study uses the publicly available MIT-BIH Arrhythmia Database [24], a benchmark dataset extensively utilized in ECG classification research. It contains 48 half-hour two-channel ambulatory ECG recordings from 47 subjects, sampled at 360 Hz with 11-bit resolution over a 10 mV range. The dataset includes expert-annotated R-peak locations and beat-type labels.

For this study, only Lead II signals were used, as it is the most commonly employed lead for arrhythmia detection and provides clinically relevant morphological information. Annotation symbols were grouped into two classes:
– Class 0 (Normal beats): N, L, and R labels
– Class 1 (Arrhythmic beats): V (ventricular ectopic), A (atrial ectopic), and F (fusion)

The original dataset exhibits severe class imbalance, with 6,279 normal beats and only 43 arrhythmic beats (Class 1).

Preprocessing was performed using the WFDB Python Toolkit [24], which enables direct access to signal waveforms and annotations. The following steps were applied:

a. Beat segmentation – each heartbeat was extracted into a 128-sample fixed-length window centered on the R-peak. This window length captures both pre- and post-R-peak morphology while keeping computational complexity manageable.
b. Z-score normalization – each segment was normalized individually by subtracting the mean and dividing by the standard deviation:

$$x' = \frac{x-\mu}{\sigma}$$

This ensures uniform amplitude scaling across beats and improves convergence during training.
c. Artifact removal – beats with incomplete annotations or excessive baseline drift (>15% of signal range) were excluded. This quality control step reduced noise-related misclassifications.

Each heartbeat was segmented into fixed-length windows comprising 128 samples, centered precisely at the R-peak to capture the most relevant part of the cardiac cycle for classification. For labeling purposes, the beats were grouped into two classes based on the annotation symbols. Class 0 included normal beats represented by N, L, and R annotations, while Class 1 consisted of arrhythmic beats categorized under V (ventricular ectopic), A (atrial ectopic), and F (fusion) types. To ensure consistency across samples and mitigate the effects of amplitude variation and baseline drift, each beat underwent z-score normalization. This normalization technique involved subtracting the mean and dividing by the standard deviation for each individual beat, thus standardizing the input distribution and improving model convergence during training.

2.2. Data augmentation

To address the extreme class imbalance and improve model generalization, two lightweight yet effective augmentation techniques were applied exclusively to Class 1 beats:

- Jittering – Gaussian white noise $N(0,\sigma^2)$ with σ in the range [0.005, 0.02] was added to the signal. This simulates physiological and acquisition-related variations in ECG morphology.
- Scaling – amplitude scaling factors between 0.9 and 1.1 were applied to mimic patient-specific variations in ECG voltage.

Augmentation was repeated iteratively until Class 1 contained the same number of beats as Class 0 (6,279 samples each), resulting in a balanced dataset of 12,558 beats. This choice of augmentation over more complex methods such as GANs or SMOTE was motivated by computational efficiency and the low risk of generating unrealistic beats.

A critical challenge encountered in the dataset was the severe class imbalance, with only 43 arrhythmic beats available compared to 6,279 normal beats. Such imbalance can severely bias the model towards the majority class, reducing its sensitivity to minority-class events. To address this, data augmentation techniques were applied exclusively to Class 1 samples. Two primary augmentation strategies were employed: jittering and scaling. Jittering involved adding small amounts of random noise to the ECG signal to simulate variability commonly observed in real-world recordings. Scaling, on the other hand, entailed amplifying the ECG beat signal to mimic variations in signal amplitude due to physiological or acquisition-related factors. These augmented samples were iteratively generated until the number of Class 1 samples matched that of Class 0, resulting in a balanced dataset with 6279 samples in each class. This augmentation process not only addressed class imbalance but also introduced variability that contributed to the generalization capability of the deep learning model.

2.3. Model architecture

Two GRU-based deep learning architectures were implemented:

2.3.1. Unidirectional recurrent neural network

In medical image segmentation tasks, particularly in areas such as tumor boundary identification or lesion segmentation, U-Net has emerged as a powerful deep learning architecture due to its encoder-decoder structure and high localization accuracy. The Figure 1 illustrates a step-by-step workflow typically used to implement a U-Net-based segmentation model.

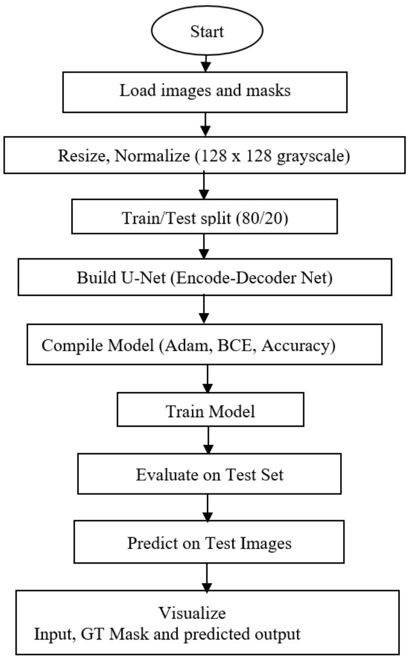

Figure 1. Workflow for training and evaluating a U-Net model for medical image segmentation

The unidirectional RNN model shown in Figure 1 employed in this study is based on the GRU, a variant of RNN known for its efficiency in capturing temporal dependencies in sequential data. The model begins with a GRU layer comprising 64 units, which processes the ECG signal in a forward direction—from the earliest to the latest time step—allowing it to learn the sequential structure of heartbeats. To prevent overfitting and improve the generalization ability of the model, a dropout layer with a rate of 0.3 is applied immediately after the GRU layer. This is followed by a dense (fully connected) layer containing 32 neurons, which transforms the temporal features into a higher-level representation. The final layer is a dense output node with a sigmoid activation function, which outputs a probability value indicating whether the input beat is normal or arrhythmic. This straightforward architecture is lightweight and effective for modeling temporal sequences, but its limitation lies in processing information in only one temporal direction.

The flowchart begins with the loading of input images and their corresponding ground truth (GT) masks, which are essential for supervised learning. These images are then resized and normalized—typically to a resolution of 128×128 in grayscale format—to ensure uniformity and reduce computational complexity. The dataset is split into training and testing subsets, commonly in an 80/20 ratio, to allow both learning and evaluation. A U-Net architecture, which consists of an encoder (for context capture) and a decoder (for precise localization), is constructed. The model is compiled using the Adam optimizer, binary cross-entropy loss (BCE), and accuracy as the evaluation metric. Training is carried out on the prepared dataset. Once training concludes, the model's performance is assessed on the test set, followed by predictions on unseen test images. Finally, results are visualized by comparing input images, GT masks, and the model's predicted outputs, enabling qualitative assessment of segmentation accuracy.

2.3.2. Bidirectional recurrent neural network

Accurate classification of ECG beats is vital in automated cardiac monitoring systems. Leveraging deep learning, particularly RNN like GRU, offers a robust approach for modeling temporal patterns in ECG signals. The Figure 2 illustrates a complete pipeline for classifying ECG beats using a GRU-based deep learning model trained on the MIT-BIH dataset.

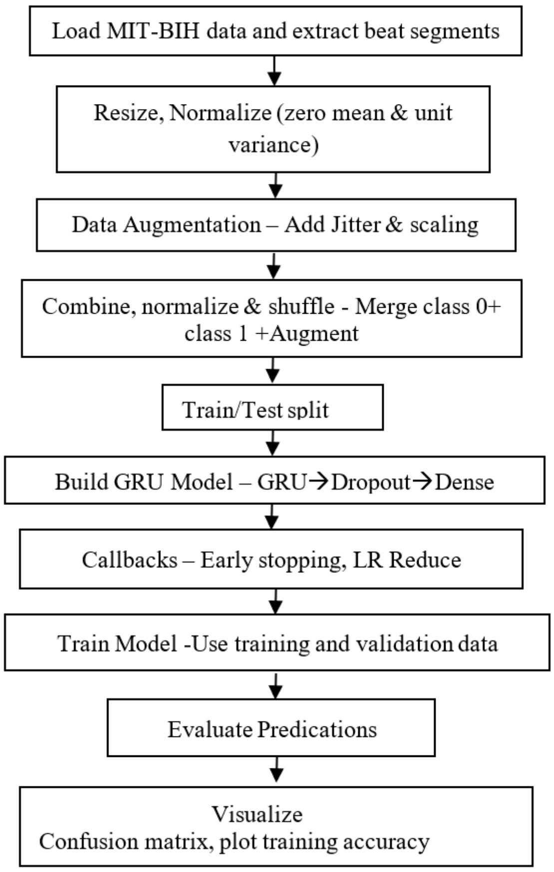

Figure 2. Workflow for ECG beat classification using a GRU-based deep learning model

The Bi-RNN model shown in Figure 2 extends the unidirectional architecture by incorporating a bidirectional GRU layer with 64 units. Unlike the unidirectional GRU, which only considers past inputs, the bidirectional GRU processes the sequence in both forward and backward directions, enabling the model to learn from both past and future contexts of each time step. This dual-pass mechanism significantly enhances the model's ability to capture complex temporal dependencies in ECG signals, which often include features that span both sides of the R-peak. Similar to the unidirectional model, a dropout layer with a 0.3 rate is used after the Bi-GRU layer to reduce overfitting. This is followed by a dense layer with 32 units that consolidates the bidirectional features, and finally, a sigmoid-activated dense output layer that classifies the beat as either normal or arrhythmic. By leveraging information from the entire sequence, the bidirectional GRU model typically offers improved accuracy and robustness in sequence classification tasks. The process begins with loading ECG recordings from the MIT-BIH dataset and extracting individual beat segments around annotated peaks.

These segments are then normalized to have zero mean and unit variance, ensuring uniformity across samples. To address class imbalance, data augmentation techniques such as jittering and amplitude scaling are applied to underrepresented classes. All beat segments—original and augmented—are merged, normalized again, and shuffled to ensure randomization. The dataset is then split into training and testing sets. A GRU-based neural network is constructed, featuring a sequence of GRU, dropout, and dense layers, optimized for temporal signal processing. Training utilizes callbacks like early stopping and learning rate reduction to prevent overfitting and stabilize learning. The model is trained using the training set and validated on the validation split. Post-training, predictions are evaluated for performance, and results are visualized using a confusion matrix and training accuracy plots to analyze classification effectiveness and training progression.

2.4. Training configuration

The training of the GRU-based ECG beat classification model was configured using the Adam optimizer, known for its efficiency and adaptive learning rate capabilities. The loss function employed was binary crossentropy, appropriate for the binary classification task. Training was conducted using a batch size of 32 samples, with the number of epochs set to a maximum of 30. However, to prevent overfitting and optimize learning, early stopping and learning rate reduction callbacks were incorporated. The dataset was partitioned into 80% for training and 20% for testing, ensuring a sufficient amount of data for both model learning and performance evaluation.

3. RESULTS AND DISCUSSION

3.1. Introduction to results

This study aimed to evaluate the effect of bidirectional temporal modeling on ECG arrhythmia detection by comparing unidirectional and bidirectional GRU architectures under balanced data conditions achieved through jittering and scaling augmentation. While both models were trained under identical conditions, the results show a substantial performance advantage for the bidirectional model, particularly in minority-class detection. This addresses a key limitation of many prior ECG classification models, which often report high overall accuracy but fail to maintain balanced sensitivity across classes

The results shown in Table 1 clearly demonstrate the superiority of the Bi-RNN over the traditional RNN across all evaluated metrics. The RNN achieved an overall accuracy of just 54.0%, with extremely poor recall (9.0%) and F1-score (17.0%) for Class 1 (typically the minority or abnormal class), indicating severe limitations in detecting critical beat types. In stark contrast, the Bi-RNN achieved an outstanding 98.17% accuracy, with high precision (98.39%), recall (97.92%), and F1-score (98.16%) for Class 1, showcasing its ability to reliably classify both normal and abnormal ECG beats.

Table 1. Performance comparison of standard RNN vs. Bi-RNN on ECG beat classification

Metric	RNN	Bi-RNN
Accuracy	54.0%	98.17%
Precision (Class 1)	95.0%	98.39%
Recall (Class 1)	9.0%	97.92%
F1-score (Class 1)	17.0%	98.16%
ROC AUC	-	0.9968
Matthews Corrcoef	-	0.9634
Cohen's Kappa	-	0.9634
log loss	-	0.0628
Balanced accuracy	54.0%	98.17%

3.2. Summary of key findings

ECG beat classification relies heavily on the ability of machine learning models to interpret sequential patterns in biomedical signals. RNN and Bi-RNNs are both designed for such time-series data, but Bi-RNNs offer the distinct advantage of processing inputs in both forward and backward directions. This enables better temporal context comprehension, which is crucial for detecting subtle abnormalities in ECG waveforms. The table below presents a comparative performance analysis of standard RNN and Bi-RNN models for ECG beat classification.

The unidirectional GRU achieved an overall accuracy of 54.0%, with precision of 95.0% but recall of only 9.0% for arrhythmic beats, resulting in an F1-score of 17.0% for Class 1. These results indicate severe under-detection of arrhythmic beats despite the model's ability to correctly classify the majority class. In contrast, the bidirectional GRU attained 98.17% accuracy, 98.39% precision, 97.92% recall, and a 98.16% F1-score for Class 1, demonstrating a well-balanced ability to detect both normal and arrhythmic beats.

Advanced metrics further highlight the robustness of the Bi-RNN model. It recorded a near-perfect ROC AUC of 0.9968, reflecting excellent discrimination between classes. The Matthews Correlation Coefficient (0.9634) and Cohen's Kappa (0.9634) indicate a strong agreement between predictions and true labels, even in the presence of class imbalance. A low log loss (0.0628) confirms the model's confident and accurate probability estimates.

The 3D horizontal bar plot shown in Figure 3 provides a comparative visualization of key performance metrics for two neural network architectures—RNN and Bi-RNN—in the context of ECG beat classification. Each metric, including accuracy, precision, recall, F1-score, and balanced accuracy, is represented along the horizontal axis, while the vertical axis distinguishes between the two models. The depth and height of the bars encode the corresponding performance values, offering a clear visual comparison. The color gradient, scaled from 1 to 5, represents the index of metrics for enhanced interpretability. The plot reveals that Bi-RNN significantly outperforms RNN across all evaluated metrics, especially in recall and F1-score, emphasizing its superior ability to capture both past and future contextual dependencies in sequential ECG data.

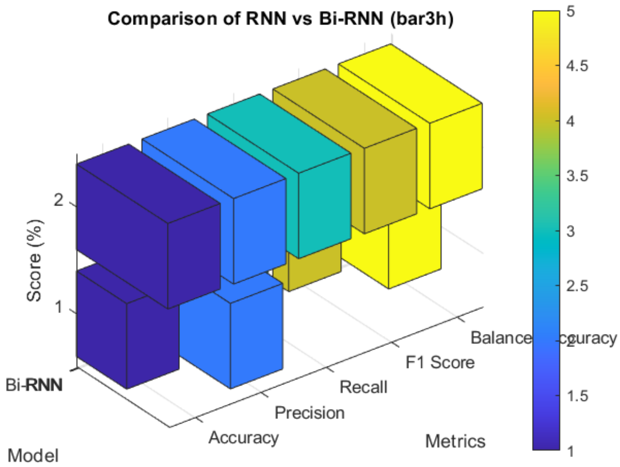

Figure 3. 3D horizontal bar plot (bar3h) comparing RNN and Bi-RNN performance metrics for ECG beat classification

3.3. Interpretation of results

The results confirm that bidirectional sequence modeling substantially enhances ECG classification performance. By processing input sequences in both forward and backward directions, the Bi-GRU model effectively leverages post-R-peak waveform information, which can be critical for detecting subtle morphological variations associated with arrhythmias. This advantage was particularly evident in the recall metric for the arrhythmic class, which improved from 9.0% in the unidirectional model to 97.92% in the bidirectional model.

This bubble chart shown in Figure 4 visually compares the performance of RNN and Bi-RNN models across five critical metrics: accuracy, precision, recall, F1-score, and balanced accuracy in ECG beat classification. Each bubble represents the score of a model for a specific metric, with its size proportional to the value. Blue bubbles indicate RNN scores, while red bubbles represent Bi-RNN scores. The chart clearly illustrates Bi-RNN's superior performance, with significantly larger bubbles across all metrics, especially in recall (97.92% vs. 9%) and F1-score (98.16% vs. 17%). The stark contrast highlights the advantage of using a bidirectional architecture, which captures both past and future sequence information more effectively than a traditional RNN. The stark performance disparity highlights the importance of capturing future context in ECG signals for effective arrhythmia detection. Unidirectional RNNs may underperform due to their limited access to contextual information that follows the R-peak, a limitation not present in Bi-RNNs. Data augmentation significantly improved the model's sensitivity to minority class features, enabling the Bi-RNN to learn robust patterns. In summary, while the standard RNN struggles with poor generalization and fails to capture temporal dependencies effectively, the Bi-RNN leverages bidirectional context to deliver highly accurate and balanced classification performance, making it the clearly superior choice for ECG beat classification tasks. These findings are consistent with Sarankumar *et al.* [23], who reported superior minority-class detection when applying Bi-GRU architectures to atrial fibrillation detection, and Lv *et al.* [22], who achieved a 4% recall improvement in Bi-LSTMs over unidirectional LSTMs for ECG classification. However, the magnitude of improvement in the present study is greater, likely due to the combination of bidirectional modeling and targeted augmentation strategies that ensured balanced exposure to minority-class examples during training.

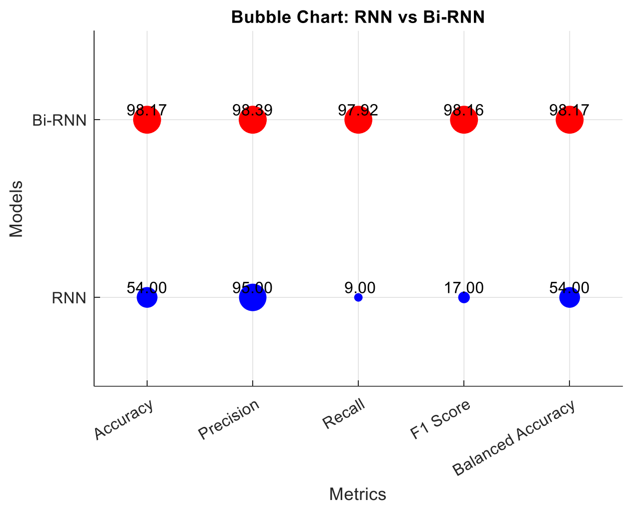

Figure 4. Bubble chart comparison of RNN and Bi-RNN models based on key performance metrics

3.4. Comparison with previous works

When compared to earlier CNN-based ECG classification models, such as Ribeiro *et al.* [12] (97% accuracy) and Ferretti *et al.* [9] (96.5% accuracy), the proposed Bi-GRU model achieves competitive or superior performance while offering better temporal feature modeling. Unlike purely convolutional approaches, which primarily capture spatial features, the Bi-GRU captures sequential dependencies that are crucial for rhythm classification. Similarly, hybrid CNN-RNN approaches like Islam *et al.* [29] achieved 98.2% accuracy, closely matching the results of this work, but often at higher computational cost due to more complex feature extraction stages.

3.5. Limitations and implications

Despite these encouraging results, it is important to note that the model was trained and evaluated on a single benchmark dataset (MIT-BIH). While the balanced augmentation strategy improved generalization within the dataset, performance on real-world clinical ECG data with different patient demographics or acquisition hardware remains to be validated. Additionally, while jittering and scaling are effective for simulating variability, they may not capture the full range of morphological patterns seen in rare arrhythmias.

The clinical implications of these findings are significant. A reduction in false negatives — as evidenced by the high recall of the Bi-GRU — directly translates to improved detection of potentially life-threatening arrhythmic events in real-time monitoring systems. The computational efficiency of the model, combined with its high accuracy, makes it a viable candidate for deployment in portable ECG devices and continuous cardiac monitoring systems.

3.6. Directions for future work

Building on these results, future research could explore integrating attention mechanisms into the Bi-GRU framework to improve interpretability, enabling clinicians to visualize which parts of the ECG sequence most influence classification decisions. Testing the model across multiple datasets, including multi-lead ECG signals, would provide stronger evidence of its generalizability. Furthermore, incorporating more advanced augmentation techniques, such as GAN-based synthetic beat generation, or combining augmentation with real beats from complementary datasets like INCART or AHA DB, could further enhance performance on rare arrhythmia classes.

4. CONCLUSION

This study presented a comprehensive comparative analysis of unidirectional and bidirectional GRU architectures for ECG arrhythmia classification using the MIT-BIH Arrhythmia Database, enhanced with jittering and scaling to address extreme class imbalance. The results demonstrated that the bidirectional GRU significantly outperformed its unidirectional counterpart across all performance metrics, achieving 98.17% accuracy, 98.39% precision, 97.92% recall, and a 98.16% F1-score. In contrast, the unidirectional GRU struggled to detect minority-class arrhythmic beats, with a recall of only 9%, highlighting the critical advantage of bidirectional temporal modeling in capturing both past and future contextual dependencies within ECG sequences. The superior performance of the Bi-GRU was further validated by advanced metrics, including an ROC-AUC of 0.9968, MCC of 0.9634, and low log loss of 0.0628, indicating high robustness and reliability even under balanced data conditions achieved via augmentation. The findings emphasize that computationally lightweight augmentation techniques, when combined with bidirectional sequence modeling, can yield highly accurate and generalizable models for clinical arrhythmia detection. Importantly, the notable reduction in false negatives suggests a tangible clinical benefit, as missed arrhythmic events can have serious implications for patient outcomes. Despite these promising results, several limitations should be acknowledged.

The study used data from a single benchmark dataset (MIT-BIH), which may not fully capture the diversity of ECG patterns across different populations, acquisition devices, or pathological conditions. Additionally, the augmentation strategies, while effective, were limited to jittering and scaling; more diverse transformations or multi-database fusion could further improve generalization. Future research could extend this work in several directions. First, evaluating the proposed approach on multiple large-scale, multi-lead ECG datasets would validate its generalizability. Second, incorporating advanced augmentation strategies such as GAN-based synthetic beat generation or hybrid oversampling methods could enhance robustness against rare arrhythmia types. Third, exploring lightweight model compression techniques, including quantization and pruning, would facilitate real-time deployment on wearable or portable ECG monitoring devices. Finally, integrating attention mechanisms into bidirectional architectures could improve interpretability, enabling clinicians to better understand the temporal features influencing model predictions. Overall, the findings of this study strongly support the use of bidirectional GRU architectures, coupled with efficient data augmentation, as a practical and high-performing solution for automated ECG arrhythmia detection, with clear potential for deployment in real-world clinical monitoring systems.

ACKNOWLEDGMENTS

The authors declare that no individuals contributed to the work in a manner that requires acknowledgment.

FUNDING INFORMATION

The authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sabura Banu Urundai Meeran	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Nafeena Abdul Munaf	✓					✓	✓			✓	✓	✓		

C	:	**C**onceptualization	I	:	**I**nvestigation	Vi	:	**Vi**sualization
M	:	**M**ethodology	R	:	**R**esources	Su	:	**Su**pervision
So	:	**So**ftware	D	:	**D**ata Curation	P	:	**P**roject administration
Va	:	**Va**lidation	O	:	Writing - **O**riginal Draft	Fu	:	**Fu**nding acquisition
Fo	:	**Fo**rmal analysis	E	:	Writing - Review & **E**diting			

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

Not applicable. No personal or patient-related data were included in this study that required informed consent.

ETHICAL APPROVAL

Not applicable. This study did not involve any human or animal subjects requiring institutional ethical review.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] World Health Organization, “Cardiovascular diseases (CVDs),” *World Health Organization*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection with deep neural networks,” *Nature Medicine*, vol. 25, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.

[3] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Ed. Patrick McSharry. Boston: Artech house, 2006.

[4] Q. Xiao *et al.*, “Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review,” *Applied Sciences*, vol. 13, no. 8, pp. 1-25, 2023, doi: 10.3390/app13084964.

[5] A. Turnip, M. I. Rizqywan, D. E. Kusumandari, M. Turnip, and P. Sihombing, “Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection,” *International Conference on Innovation in Education, Science and Culture (ICIESC-2017)*, Medan, Indonesia, 2018, vol. 970, doi: 10.1088/1742-6596/970/1/012012.

[6] S. Faziludeen and P. Sankaran, “ECG Beat Classification Using Evidential K -Nearest Neighbours,” *Procedia Computer Science*, vol. 89, pp. 499-505, 2016, doi: 10.1016/j.procs.2016.06.106.

[7] 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, Jul. 2004, doi: 10.1109/TBME.2004.827359.

[8] F. Khan, X. Yu, Z. Yuan, and A. ur Rehman, “ECG classification using 1-D convolutional deep residual neural network,” *PLoS ONE*, vol. 18, no. 4, pp. 1-22, 2023, doi: 10.1371/journal.pone.0284791.

[9] J. Ferretti, V. Randazzo, G. Cirrincione, and E. Pasero, “1-D Convolutional Neural Network for ECG Arrhythmia Classification,” *Progresses in Artificial Intelligence and Neural Systems*, Singapore: Springer Singapore, pp. 269-279, Jan 2021, doi: 10.1007/978-981-15-5093-5_25.

[10] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, “ECG heartbeat classification: A deep transferable representation,” in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, New York, NY, USA, 2018, pp. 443-444, doi: 10.1109/ICHI.2018.00092.

[11] G. Sannino and G. De Pietro, “A deep learning approach for ECG-based heartbeat classification for arrhythmia detection,” *Future Generation Computer Systems*, vol. 86, pp. 446-455, 2018, doi: 10.1016/j.future.2018.03.057.

[12] A. H. Ribeiro *et al.*, “Automatic diagnosis of the 12-lead ECG using a deep neural network,” *Nature Communications*, vol. 11, 2020, doi: 10.1038/s41467-020-15432-4.

[13] M. Wang *et al.*, “Multi-task learning for multi-label electrocardiogram disease detection via point-level delineation-guided fusion,” *Engineering Applications of Artificial Intelligence*, vol. 160, p. 111894, 2025, doi: 10.1016/j.engappai.2025.111894.

[14] 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, Aug. 2018, doi: 10.1016/j.compbiomed.2018.03.016.

[15] G. Kothai, R. S. S. Mithran, B. K. Balaji, S. Suryaprakash, M. Sanjai, and P. Sanjai, “ECG Arrhythmia Classification Using CNN-BiLSTM Models for Enhanced Diagnostic Accuracy,” in *2024 International Conference on Computing and Intelligent Reality Technologies (ICCIRT)*, Coimbatore, India, 2024, pp. 326-330, doi: 10.1109/ICCIRT59484.2024.10922026.

[16] T. Lakshman, S. S. Kumar, U. S. Kumar, and Y. S. Sekhar, “Efficient Deep Learning Model for ECG Signal Classification: A Micro-Class Approach,” *International Journal of Novel Trends and Innovation*, vol. 2, no. 3, pp. a56-a61, Mar. 2024.

[17] M.-R. Zhu, J.-D. Liu, and J.-Z. Ji, “Electrocardiogram Signal Classification Based on Bidirectional LSTM and Multi-Task Temporal Attention,” *Journal of Computer Science and Technology*, vol. 40, pp. 1401–1413, 2025, doi: 10.1007/s11390-025-4330-6.

[18] S. Śmigiel, K. Pałczyński, and D. Ledziński, “ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset,” *Entropy*, vol. 23, no. 9, pp. 1-20, 2021, doi: 10.3390/e23091121.

[19] E. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification," *IEEE Access*, vol. 9, pp. 103452-103464, 2021, doi: 10.1109/ACCESS.2021.3098986.

[20] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[21] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing,* Vancouver, BC, Canada, 2013, pp. 6645-6649, doi: 10.1109/ICASSP.2013.6638947.

[22] Q. J. Lv *et al*., "A Multi-Task Group Bi-LSTM Networks Application on Electrocardiogram Classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1-11, 2020, doi: 10.1109/JTEHM.2019.2952610.

[23] R. Sarankumar, M. Ramkumar, K. Vijaipriya, and R. Velselvi, "Bidirectional gated recurrent unit with auto encoders for detecting arrhythmia using ECG data," *Knowledge-Based Systems*, vol. 294, Jun. 2024, doi: 10.1016/j.knosys.2024.111696

[24] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, May-Jun. 2001, doi: 10.1109/51.932724.

[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[26] O. Düzyel and M. Kuntalp, "Data Augmentation with GAN increases the Performance of Arrhythmia Classification for an unbalanced dataset,*" arXiv*, 2023, doi: 10.48550/arXiv.2302.13855.

[27] Q. Pan, X. Li, and L. Fang, "Data Augmentation for Deep Learning-Based ECG Analysis," *Feature Engineering and Computational Intelligence in ECG Monitoring.* 2020, PP. 91-111, doi : 10.1007/978-981-15-3824-7_6.

[28] C. Zhao, B. Lai, Y. Xu, Y. Wang, and H. Dong, "MAK-Net: A Multi-Scale Attentive Kolmogorov-Arnold Network with BiGRU for Imbalanced ECG Arrhythmia Classification," *Sensors (Basel*), vol. 25, no. 13, pp. 1-19, 2025, doi: 10.3390/s25133928.

[29] M. S. Islam, M. N. Islam, N. Hashim, M. Rashid, B. S. Bari, and F. A. Farid, "New Hybrid Deep Learning Approach Using BiGRU-BiLSTM and Multilayered Dilated CNN to Detect Arrhythmia," *IEEE Access*, vol. 10, pp. 58081-58096, 2022, doi: 10.1109/ACCESS.2022.3178710.

BIOGRAPHIES OF AUTHORS

Dr. Sabura Banu Urundai Meeran was born in Chennai, Tamil Nadu, India on May 8, 1975. She is presently working as a Professor in the Department of Electrical and Electronics Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India. She has an illustrious academic and professional background in engineering, innovation, and consultancy. She completed her Ph.D. in Intelligent Modeling and Control at College of Engineering, Guindy, Anna University (2006–2009) after earning her M.S. by Research in the same Institution and B.E. in Instrumentation and Control Engineering. She also holds a Diploma in Computer Science Engineering and a B.Sc. in Mathematics. With over two decades of experience, she has served in various leadership roles, including Director at Arcres Techsol Pvt. Ltd., Founder & CEO, Intellectz Tech for All and Head of Departments at esteemed institutions like SRM Institute of Science and Technology and B.S. Abdur Rahman University. She has mentored numerous student projects that garnered national and international accolades, including awards, patents, and startup funding. Her notable achievements include 34 international journal publications in leading Journals like Taylor and Francis, Applied Soft Computing, ACTA Press, AMSE, and Inderscience. Contributions to five book chapters, and 132 conference presentations. She can be contacted at email: saburabanuu@saveetha.ac.in and saburasec@gmail.com.

Dr. Nafeena Abdul Munaf is Assistant Professor and Coordinator for Electrical and Electronics Engineering program. Villa College, Maldives. Besides that, she has published many papers in Scopus, Web of Science, Science Indexed and UGC journals. Her recent published paper in IEEE got 49 citations on the title' An ML approach for Household, power consumption. Currently she is a reviewer in reputed international and national journals and editor in special issues of conference publications. Also, she has given many keynote lectures in reputed colleges and universities in Smart Grid Challenges and Towards Sustainable University Residences: Comprehensive Energy Audit, Retrofit Techniques and Financial Implications Sustainability Issues in Environment. She is interested in writing articles and publishing on Life Cycle Analysis of various products and Smart Energy management systems for smart grid. Also, she has received grants on SDG based researches. She can be contacted at email: nafeena.abdul@villacollege.edu.mv.