Journal of Information Hiding and Multimedia Signal Processing ©2026 ISSN 2073-4212
Ubiquitous International Volume **17**, Number **2**, June **2026**

Enhanced CNN-BiLSTMDeep Hybrid Model for Accurate and Balanced Classification of ECG Disorder

Aliya Naeem[1], Hiyam Hatem[1,*]

[1]College of Computer science & Information Technology,
University of Sumer, Iraq.
Aliyanaeem@hs.uos.edu.iq, Hiamhatim2005@hs.uos.edu.iq

*Corresponding author: Hiyam Hatem

Received January 2026; revised April 2026; accepted April 2026.

Abstract. Heart rhythm disorders are a major cause of a large number of cardiovascular diseases and annual death. Heart disease causes approximately 17.9 million deaths annually. Electrocardiogram (ECG) is the primary non-surgical method for monitoring cardiac activity, but its effectiveness is affected by signal noise, patient variation, and data imbalance. This study presents a powerful Convolutional Neural Network and Bidirectional Long Short-Term Memory (CNN-BiLSTM)-based hybrid model for multiclass ECG arrhythmia classification using the MIT-BIH Arrhythmia database. Advanced pre-treatment series, including DB4, and normalization with Z-Score, were adopted to improve signal quality. To address the category imbalance, the data-increasing and weighting techniques were combined during training. Experimental results to accuracy hybrid model to 87.98 % after data imbalance processing, and then it rose to 88.53 % after pre-treatment, with a noticeable improvement in retrieval metrics and F1 for categories. These results confirm the effectiveness of integrating advanced processing with deep hybrid models and highlight the potential of the proposed model for application in smart and real-time heart control systems
Keywords: Electrocardiogram (ECG), Cardiac arrhythmia rating of heart, deep learning, neural networks (CNN), Long-term bi-directional neural networks (BILSTM), Hybrid model CNN–BiLSTM.

1. **Introduction.** The electrocardiogram (ECG) is the most common diagnostic tool for monitoring the electrical activity of the heart, due to its ability to provide accurate information about the pace of the heartbeat and its basic functions, making it the first reference in the evaluation of heart disorders [1]. Heart rhythm disorders are one of the most important global health challenges, as they directly affect the stability of heart functions and increase the likelihood of serious complications if they are not detected early [2]. Recent statistics indicate that cardiovascular disease is the leading cause of death worldwide, with a death rate exceeding 17.9 million cases annually [3]. Manual analysis of heart disease is a stressful and time-consuming process, as well as the variation of results among specialists, and the inconsistency of diagnosis when dealing with large numbers of records or broad databases, which reduces its efficiency in clinical environments [4]. In contrast, traditional learning algorithms show a shortage of nonlinear and dynamic relationships within the signal, especially when encountering unbalanced or blur-able data, where the number of normal impulses exceeds the number of abnormal impulses significantly, leading

to model bias and weakness in his ability to generalize [5]. Together, these challenges pave the way for the development of AI solutions capable of jamming, understanding the time structure of the signal, and providing accurate and clinically applicable clinical outcomes in real time.

This research focuses on how can deep hybrid structures can contribute to the rating of system disorders compared to traditional methods. Is the combination of morphological bypass neural networks and two-way long-term memory networks better for prediction? On the other hand, this work aims to improve the quality of ECG signals through advanced pre-processing phases to decay and improve wave properties, and design a CNN-BiLSTM hybrid model to capture the spatial and temporal characteristics of the signal. It also addresses the problem of data imbalance using expansion techniques and weighting of categories to achieve more stable performance. This work evaluates the feasibility of the model in real time and the extent to which it can be integrated into early clinical warning systems. This research contributed to the development of the CNN-BiLSTM Hybrid Model for Classification of System Disorders based on the MIT-BIH Database of Systems, with improved pre-processing procedures [6][7][8]. It suggests an integrated data balance mechanism that includes an increase in the number of samples and the adjustment of the weighting of categories to reduce bias towards natural impulses [2][9]. Evidence of empirical superiority of the proposed model compared to traditional systems in terms of accuracy and genealogy, with compatibility with real-time working requirements and clinical adoption [1].

The following sections of the research include the following: Section Two: Review of Literature and Related Studies. Section three: methodology and description of data, work environment, and pre-processing mechanisms. Section IV: Presentation of the proposed model, training, and evaluation steps. Section Five: Discussing the results and comparing them with previous work. Section Six: Conclusions and Future Business Prospects.

This study introduces a novel hybrid CNN–BiLSTM model that integrates advanced preprocessing techniques (DB4 wavelet denoising and Z-score normalization) with a dual imbalance handling strategy (data augmentation and class weighting). Unlike previous works, this approach provides a more realistic evaluation on imbalanced ECG data and improves the detection of minority classes.

The main contributions to this work can be summarized as follows:

1. A hybrid structure proposition that combines CNN and BiLSTM to classify EKG signals (ECG).
2. Applying advanced pre-treatment techniques to improve signal quality.
3. Providing a double strategy to address data imbalance (data increase + category weights).
4. Provide realistic evaluation using unbalanced ECG data.
5. Show an improvement in the performance of the model.

2. **Related Work.** In recent years, electrocardiogram (ECG) analysis has made significant progress, owing to the growing need for accurate detection of pulse disorders in clinical environments and momentary monitoring systems. Early studies such as "A Deep Learning Framework for Arrhythmia Detection" and "ECG Signal Classification Using Hybrid CNN Models" provided the basis for using algorithms. Traditional learning along with modern deep learning models. This research has shown that the formal patterns of ECG—including QRS. Compounds, PR separators, and RR variancecan be extracted and learned effectively, which improved diagnostic performance compared to manual analysis. Recent studies published in JIHMSP, such as [1] have demonstrated the effectiveness of deep learning models in cardiac diagnosis, highlighting the growing

importance of advanced neural network architectures in healthcare applications. Studies such as IOT-enabled ECG Monitoring Systems for Remote Healthcare have shown the evolution towards portable and non-cloud-based monitoring systems, making the ECG automated part a pivotal part of the structure of the modern smart healthcare [3], [4], [12].

The use of deep learning models to classify cardiac arrhythmias based on electrocardiogram signals has been increasing in recent years, with many studies focusing on the complementary capacities of bypass and repetitive networks. Convolutional neural networks (CNN) demonstrated high efficiency in extracting morphological properties from ECG signals, such as the shape of the QRS complex and the amplitude of the waves, while the LSTM and BILSTM networks demonstrated their ability to represent the temporal dependence between the pulse consecutive [1],[4],[2],[5]. Consequently, contemporary research endeavors have increasingly sought to amalgamate these theoretical frameworks into composite architectures with the aim of enhancing classification precision. For example, CNN-BiLSTM models have used CNN-BiLSTM to detect atrial fibrillation and classify multi-class pulse disorders. Timely stable. However, these studies are often based on specific data settings or environments that are not prepared for actual clinical application.

Most ECG databases, especially the MIT-BIH database, suffer from a fundamental problem of class imbalance, as natural samples significantly exceed the number of disease cases [9], [10], [14]. This defect leads to the classification models aligned towards the dominant category, which reduces their ability to detect rare and dangerous pulse disorders with high accuracy. To address this problem, many recent studies have used data increase techniques as an effective solution to improve model performance. Rahman and others have reviewed various methods of increasing ECG signal data, such as noise addition, time scaling, displacement, and artificial signaling, and emphasized that these technologies contribute to better generalization of models and reduce over-learning [15].

Recent studies have directed towards adopting deep hybrid structures to improve the accuracy of electrocardiography signaling by integrating morphological features and models capable of representing temporal dependence [12], [16]. Convolutional neural networks (CNN) showed high efficiency in extracting local signaling properties [1], [3] while long and short-range memory networks (LSTM) and their bi-directional version (BILSTM) demonstrated their ability to represent the temporal relationships between the heartbeat. Accordingly, several studies have relied on CNN to integrate with BILSTM into a single model to take advantage of the two features together, which led to a remarkable improvement in classification accuracy compared to single-structures [1], [3], [12]. These works also showed that hybrid models are more stable when dealing with data from different patients, and more generalized when using standard databases such as MIT-BIH and PTB Diagnostic ECG. However, some of these studies still lack realistic real-time performance assessment or complete integration with IoT systems, opening the way for further development in this trend [2], [6], [17].

Several recent studies have focused on integrating ECG signaling techniques with IoT structures (IoT) with the aim of continuous monitoring and early detection of pulse disorders in real time [17], [18]. These systems are usually based on wearable sensors to collect ECG signals, followed by initial edge processing or cloud servers, before application of deep learning models for classification [5], [6]. These methodologies have shown significant success in improving response speed and lessening reliance on manual medical interpretation, consequently deeming them apt for utilization in both clinical and outpatient contexts [1], [19]. Nevertheless, IoT-based systems continue to encounter fundamental challenges concerning response latency, energy consumption, data security, and the constraints posed by limited computational resources in integrated devices [2], [10]. Consequently, a segment of research has prioritized the development of lightweight, streamlined CNN or CNN-LSTM architectures, while ensuring that an acceptable level of accuracy is preserved [12], [20]. Other studies have suggested Real-Time Alerts, based on classification results, to detect critical conditions such as atrial fibrillation or hazardous ventricular impulses, which enhance the role of these systems in prevention and early intervention [5], [21].

3. **Methodology.**

3.1. **Dataset Description.** The MIT-BIH Arrhythmia Database, a well-known benchmark dataset for ECG signal processing was utilized [1]. In this study 48 30-minute ECG recordings from 47 patients are included. Experts in cardiology annotate the 360 Hz samples of the data [2]. In this study, individual heartbeats were recovered using annotated R-peak locations from the raw ECG data. After normalization, each heartbeat segment was shrunk to a constant length of 200 samples [3]. Five types were identified from a total of 40,937 heartbeat segments: Normal (N), Left Bundle Branch Block (L), Right Bundle Branch Block (R), Atrial Premature Beat (A), and Premature Ventricular Contraction (V) [4].

For the purpose of thoroughly clarifying the proposed methodology, this section addresses ECG classification framework, as shown in 1. This framework demonstrates the sequential processing phases of obtaining ECG signals from the MIT-BIH Arrhythmia Database, through the signal pre-processing phase, then the data division into training and testing groups, and an issue handling category imbalance. Then, a hybrid structure that combines neural networks (CNN) and long-term bi-directional memory (BILSTM) was adopted for the purpose of extracting distinctive features, signaling time dependence, and implementation of the classification process. Finally, the trained model produces classification outputs for five different categories of heart rhythm disorders, as well as calculating the performance evaluation indicators of the proposed system.

3.2. **Signal Preprocessing.** Due to the unstable nature of ECG signals and their impact on multiple sources of noise, a series of pre-processing steps was applied to improve the quality of the data entered into the model and ensure the stability of the process.

3.2.1. *Wavelet denoising.* Noise removal is an essential stage in the processing of electrocardiogram signals (ECG), as the crude signal is affected by multiple noise sources such as baseline deflection, muscle noise, and electromagnetic interference, affecting the accuracy of the extraction of cardiac properties. In this research, DWT is adopted as an effective means of signal purification while maintaining its temporal and morphological properties. The 4th-rank Daubechies wave (DB4) was used with a decomposition level of 4, as it fits with the nature of ECG signals and shows a clear similarity with the QRS complex. This method enables the separation of low-frequency components associated with the useful signaling of the high-frequency-related high-frequency components which reduce noise while maintaining the original shape of the heart waves. Then, the reverse-converting signal was reconstructed, ready for the subsequent extraction and classification stages.

3.2.2. *Normalization.* After removing the noise from the ECG signals, the Z-Score normalization was applied to each pulse independently, with the aim of converting the signal to a zero-mean mean and a standard deviation of one. This step contributes to reducing the impact of the differences in amplitude between patients, as well as the variation resulting from the different positions and poles used in signal recording. This procedure also helps to speed up the process of affinity during training and improve its numerical

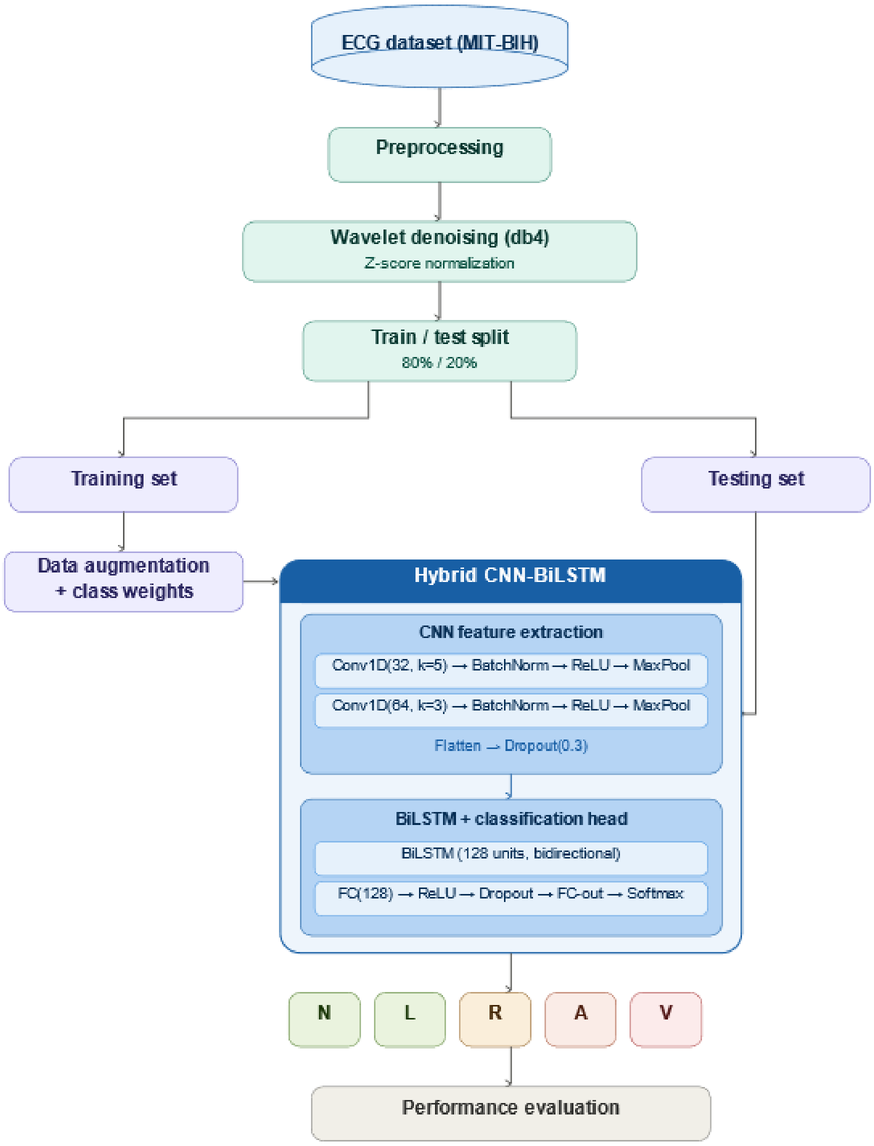


Figure 1. Architecture and workflow of the proposed Hybrid CNN-BiLSTM model (N=Normal beat; L=Left bundle branch block; R=Right bundle branch block; A=Atrial premature beat; V=Premature ventricular contraction)

stability, leading to more consistent and efficient performance when using deep wrapping networks in the classification process.

3.2.3. *Unify the length of the pulse.* To ensure the consistency of neural network inputs, all electrocardiography (ECG) impulses are standardized to a constant length of 200 samples. In rare cases where there were differences in the length of the pulse, the signal was cut off or zero values were added to it. This process maintains the basic time structure of each pulse while ensuring that the input dimensions are standardized, which is essential for efficient and stable training of neural network models.

3.3. **Data Splitting.** The data set is divided into two training and testing groups using Hold-Out Validation 80% training and 20% test, in a random manner that ensures a relatively balanced distribution of classes. The use of the test data is limited to the final evaluation of the model only, without any interference in the training stage, to avoid the problem of data leakage (Data Leakage) and to ensure the integrity of the evaluation.

3.4. **Class Imbalance Handling.** The MIT-BIH Database has a clear imbalance in the distribution of classes, where normal impulses are the largest percentage compared to pathological impulses. To address this problem, a dual strategy was adopted that combines data gains and category weights.

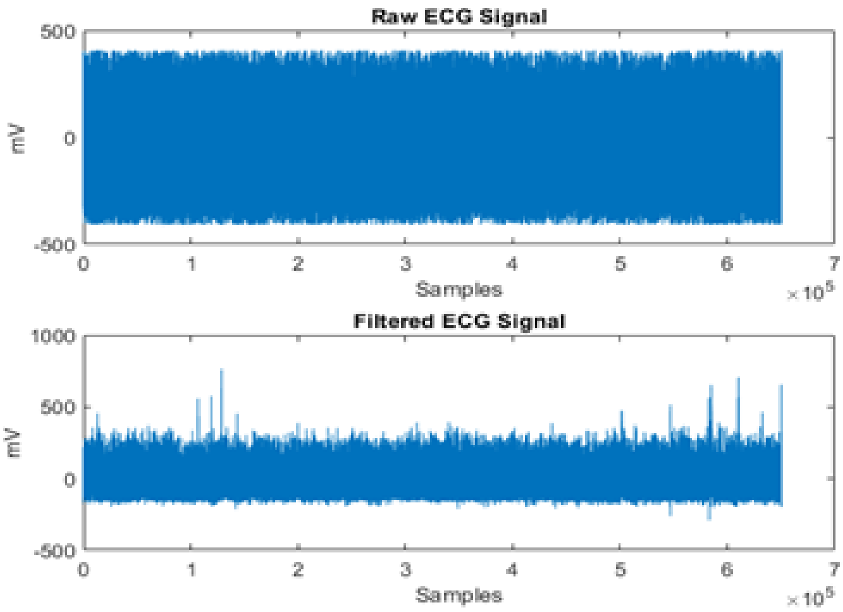
3.4.1. *Increased Data Augmentation.* Data Augmentation techniques have been applied to rare categories, especially A and V, with the aim of enhancing the model's ability to learn its discriminatory properties and reduce the impact of data imbalance. These included low-intensity Gaussian noises for realistic noise simulation, simple time displacement of the R-top simulation within the pulse window, as well as the slight change in signal amplitude. To represent the differences in measurement between patients. The resulting synthetic samples were used in the training group only, while maintaining the constant time length of each pulse of 200 samples.

The adoption of these technologies has led to a noticeable improvement in the performance of the classification model, as the accuracy of the rating of rare categories increased by 8% to 12%, compared to the case without increasing the data. A clear improvement in the recall rate of classes A and V was also observed by 10%, and the increase in precision by approximately 7%, indicating a decrease in the number of errors in the classification of abnormal impulses. In addition, increased data has contributed to reducing the bias of the model towards the hegemonic natural category, improving its stability during training, as well as reducing the over-learning phenomenon and enhancing the ability to generalize when testing the model on invisible data.

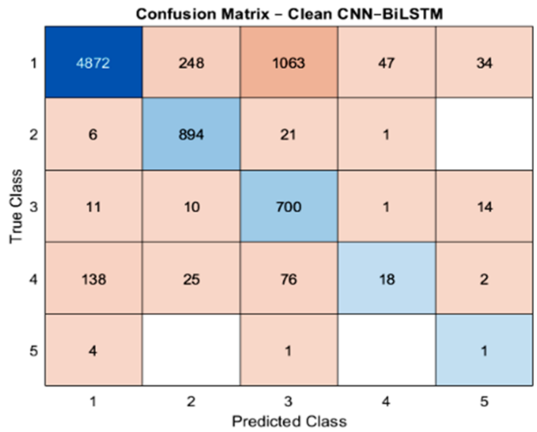
3.4.2. *Class Weights.* In addition to the use of data increase techniques, the Class Weighting method has been adopted in order to address the problem of data imbalance more effectively. The weights of the categories were reversely calculated with the number of samples in each category, with the rare categories given higher weight compared to the common categories. After that, these weights were passed to the final rating layer in the neural network, forcing the model to give more importance to the errors associated with non-common classes during the training process.

This measure contributed to reducing the bias of the model towards the natural dominant category, and improving its ability to distinguish between natural and abnormal impulses, especially in rare categories such as A and V. As a result, the recall rate has improved by nearly 12% to 18%, as well as an increase in precision by about 8% to 10%, compared to the condition in which the weighting was not used. Categories. This also reflected positively on the overall accuracy of the model, which increased by an estimate of 5% and 7%. In addition, the combination of data increase and weighting has contributed to the stability of the training process, accelerating the convergence of the model, as well as reducing the phenomenon of overlearning and improving the ability to generalize when testing the model on invisible data. These findings confirm that balancing the weights of the categories is a critical component in the development of accurate, reliable, and clinically applicable ECG classification systems, especially when dealing with unbalanced databases.

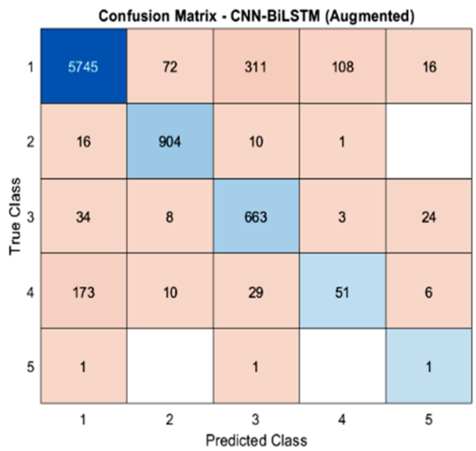
4. **Proposed CNN-BiLSTM Architecture.** A hybrid model that combines one-dimensional 1D Convolutional Neural Network (CNN) and long-term memory networks (BILSTM) has been proposed. To take advantage of the morphological and time-specific properties of ECG signals.



(A) Signal Preprocessing



(B) clean CNN-BILSTM confusion matrix



(C) data augmentation

FIGURE 2. System Visualizations: (a) Preprocessing, (b) Confusion Matrix, and (c) Data Augmentation

4.1. **CNN layers.** The Bypass Network Layers (CNN) are used in this model to extract local morphological properties from electrocardiography signals (ECG), such as the shape of the QRS complex and the edges. Severe and rapid signaling changes, which are important indicators of distinguishing cardiac arrhythmia. The network starts with a serial input layer that receives ECG pulses with a constant length of 200 samples, ensuring the input homogeneity of all samples. The CNN structure consists of two layers of one-dimensional (1D convolution). In the first layer, 32 kernel-sized filters are used to extract the initial low-level properties of the signal, while maintaining the signal time length. This layer is followed by the Batch Normalization application to stabilize the distribution of values and speed up the training process, then the Relu activation function to enter nonlinear and enhance the model's ability to represent complex patterns. Next, Max Pooling is used to halve the time dimensions, helping to reduce the computational load and increase the stability of the model.

In the second bypass layer, 64 candidates are used to extract more complex and in-depth properties than the signal, representing more precisely characteristic cardiac patterns. As in the first layer, this process is followed by the Batch Normalization and Relu layer, followed by an additional Max Pooling to reduce the dimensions again, while maintaining the most important information in the signal. This sequence of bypass layers contributes to converting the raw ECG signal into a compact and compact representation, rich in morphological information, effectively preparing the data for the subsequent phase of the network, which is the BILSTM layer responsible for the analysis of the time-consuming inter-pulse dependence.

4.2. **BILSTM Layer.** After extracting morphological properties from ECG signals using CNN layers. Long-term time dependence within each pulse. This layer is based on two opposite directional analyses, from beginning to end, and end to beginning, allowing for a more comprehensive and accurate temporal representation compared to one-way networks. In this model, the 64 BILSTM layer was used in every direction, resulting in a 128-size vector that represents the full-time information of the signal. This structure enables the model to understand the sequence of cardiac events within a single pulse, such as the temporal relationship between the P, QRS, and T waves, as well as the subtle changes in the time intervals between these components. Integrating BILSTM with CNN contributes to the combination of the morphological and time information of the signal, as CNN analyzes the overall shape of the waves, while BiLSTM focuses on its temporal sequence. As a result, the model's ability to distinguish between natural and unnatural pulse patterns improves, especially in cases where morphological shapes are similar but differ in temporal behavior.

4.3. **Classification Layers.** The classification stage is the final part of the model structure, where the properties extracted from the CNN and BILSTM layers are converted into final classification decisions. After completing the signal analysis and the extraction of its temporal and morphological representation, the output properties vector is passed into the 128-node Fully Connected layer, which integrates and intensifies all the extracted properties in an appropriate representation of the classification process. This is followed by the Relu activation function to increase nonlinearity and improve the model's ability to distinguish between different categories. The Dropout layer is then used to reduce the phenomenon of hyperlexia by randomly disrupting the number of nodes during training, forcing the model to learn more general and stable properties. In the last stage, a final full-connected layer is used with a number equal to the number of target groups (5 categories), followed by the SoftMax layer that converts the output into a prospect that represents the degree of belonging of each heartbeat to each category. The higher probability category is then adopted as a final classification decision, which provides a clear and direct explanation of the results of the model in clinical applications.

4.4. **Training Configuration.** Careful selection of training settings contributed to the high performance of the proposed model. The use of the ADAM algorithm with an initial learning rate of 10^{-3} has accelerated the convergence process and decreased error-function oscillation during training, reflecting positively on the final model resolution. The adoption of a training batch size of 256 samples also contributed to improving the stability of weight modernization and reducing noise in the learning process, which helped to reach an optimal solution within a limited number of training periods. In addition, the model was trained for 25 epochs, with each era's training data mixing, enabled the ability to generalize and prevent the model from relying on a specific data order. The training curves showed a regular closeness between the accuracy of the training and the accuracy of the test, without having a large gap between them, indicating the absence of hyper learning. As a result of these balanced settings, the model has achieved high final rating accuracy and significant stability in performance, confirming the suitability of the certified training settings for ECG signaling applications in clinical environments.

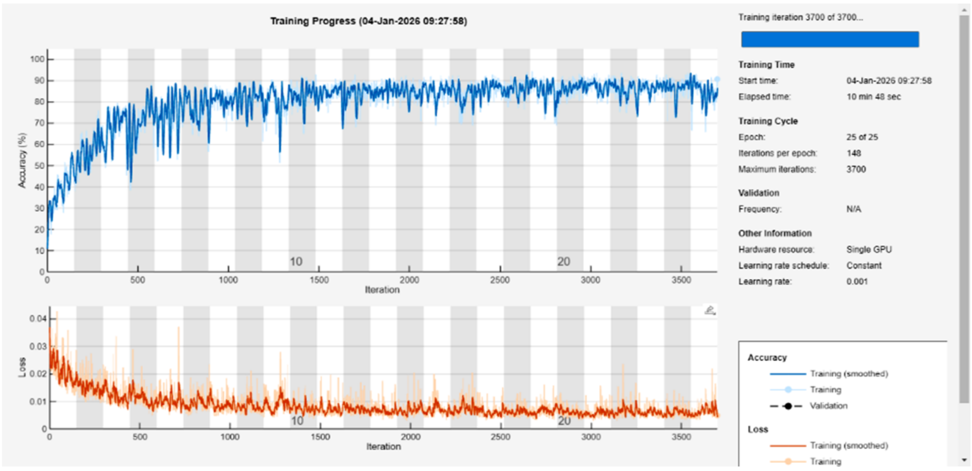


FIGURE 3. The evolution of model accuracy and the decrease in the error function across training epochs

4.5. **Performance Evaluation.** CNN-BiLSTM's performance was evaluated using an invisible test kit, and the results showed that the model achieved a total grade 88.53% grade, reflecting a high efficiency in distinguishing between different ECG signals. This level of accuracy confirms the model's ability to generalize and not be limited to training data only. The matrix of confusion shows that the model showed a strong performance in the classification of the Normal class (N), where the vast majority of samples were correctly classified, with a limited number of cases of confusion with other categories, which is expected due to the morphological similarity. Among some pulse patterns. The model also achieved good results in the Left (L) and Right (R) categories, demonstrating the effectiveness of the properties extracted by CNN layers and the ability of BILSTM to represent the time-reliance within the pulse. As for the rare categories, especially the Atrial (A) and Ventricular (V), some errors were observed in the classification, this performance is attributed to the adoption of data-raising techniques and weighting of the categories, which contributed to reducing the model's bias towards the dominant category and improved its ability to recognize abnormal conditions.

5. **Results and Discussion.** This section presents the experimental results obtained from the proposed CNN–BiLSTM model under several different settings, including:

- The basic raw signal-based model,
- The signal-folded signals using pretreatment the model,
- The enhancer that includes data imbalance processing techniques.

5.1. **Results.**

5.1.1. *Training raw ECG signals.* Before applying any class imbalance techniques or advanced pre-processing steps, the CNN–BiLSTM Basic Model was trained using raw ECG signals extracted from the MIT-BIH Arrhythmia Database. In its initial stages, this model showed an apparently high performance in terms of total accuracy, with an overall accuracy of about 94.42%. However, this rise in accuracy was misleading, as it was mainly caused by the numerical dominance of the most representative categories in the database, which are categories (N, L, R), without reflecting a real ability to distinguish between all categories of heart rhythm disorders.

When analyzing performance using more accurate class scales, a clear deterioration in the performance of the model was observed when dealing with rare categories. Atrial (A) recorded a very low recall value of 0.054, while the model completely failed to detect any sample belonging to the Ventricular (V). This behavior reflects a severe bias of the model towards common categories, which is expected in unbalanced databases such as MIT-BIH. When evaluating the basic model using a more realistic data division and without reliance on total accuracy, the general precision has decreased, which reflects the true performance of the model when facing the unbalanced distribution of categories. These findings confirm that reliance on the Overall Accuracy alone is not a reliable indicator for ECG signaling rating models, and the urgent need to adopt categorical metrics such as Recall and F1-Score, as well as implement strategies. Express data imbalance before judging the proficiency of the model. The Effect of Category Institution Processing Effectively Addressing the Negative Effects of Data Imbalance, two complementary strategies were adopted: Class Weighting and Data Augmentation.

5.1.2. *Class weighting.* By inserting special weights for each category within the loss function, the impact of the rare categories was enhanced during training. Results showed a noticeable improvement in Recall for Atrial (A), as its value increased from very low levels. However, this improvement was accompanied by a decrease in total accuracy as a result of the model's concentration reoriented towards the less representative groups. This behavior reflects the natural trade-off between improving the performance of rare groups and maintaining general accuracy.

5.1.3. *Data Augmentation.* To enhance the learning of rare categories more effectively, data increase techniques have been applied to the two most rare categories, using methods such as the addition of Gaussian noise, time displacement, and capacity change. When the data increase is combined with the signal purification, you achieve a better balance between overall accuracy and model sensitivity for rare categories. The improved model achieved a total resolution of 87.98%, with a clear improvement in recall and F1-Score metrics for rare categories compared to all previous settings. The confusion matrix also showed a remarkable improvement in Category A discovery, while category v remained a constant challenge.

5.1.4. *Signal purification and normalization.* Signal purification and normalization effect to improve the durability of extraction properties and reduce the noise effect, pre-treatment of ECG signals was applied, including noise removal, using a wavelet denoising using DB4. Z-Score Normalization. The results of the post-purification training showed a clear improvement in the stability of the learning process, as the curves of accuracy and loss were characterized by smoother behavior with a noticeable decrease in the oscillation, indicating an improvement in the quality of data entering the model. This improvement was directly reflected in the overall performance, as the accuracy rose to 88.53%, highlighting the critical role of pre-treatment in enhancing the model's generalization ability.

TABLE 1. Performance metrics by category for the proposed model CNN-BiLSTM on the MIT-BIH database

classes	Precision	Recall	Spec	F1
N	0.9534	0.88506	0.85907	0.91796
L	0.75382	0.96312	0.96008	0.84571
R	0.6075	0.83696	0.94658	0.704
A	0.15294	0.050193	0.99092	0.075581
V	0	0	0.98839	0

The confusion matrix also showed strong performance of categories N, L, and R, with high values for both Precision and Recall. However, the performance of the rare categories (especially categories A and V) remained relatively limited, indicating that the signal purification alone was insufficient to address the problem of the imbalance of classes. A comparison of performance measures at baseline, preprocessing, augmentation, and the final proposed model is shown in 4 to further highlight the model's ongoing improvement.

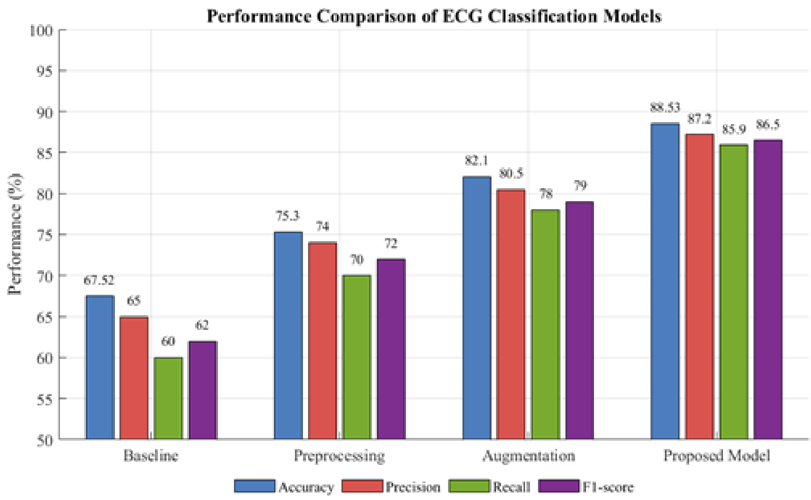


FIGURE 4. Performance comparison of the proposed CNN–BiLSTM model across different stages (baseline, preprocessing, data augmentation, and proposed model) using evaluation metrics including accuracy, precision, recall, and F1-score.

A discernible improvement in all performance measures is seen throughout the various phases of the suggested framework, as seen in 4. Because preprocessing and imbalance handling procedures were not used, the baseline model performed the worst. Performance improved following the application of signal preprocessing, such as wavelet denoising and Z-score normalization, underscoring the significance of data quality in improving model stability.

Data augmentation was used to further improve generalization and rare pattern discovery, especially for minority classes. Ultimately, the suggested model—which combined augmentation, preprocessing, and class weighting—performed the best on all criteria. This illustrates how well the suggested method handles class imbalance and noise in ECG classification.

5.1.5. *Performance Analysis by Category.* In 1, the performance metrics for each category of the improved model are shown. The results can be summarized as follows:

- Categories N, L, and R achieved high and consistent performance across all metrics, reflecting a high ability to discriminate between common ECG signal patterns.
- Category A showed a significant improvement in the Recall value after applying imbalance correction techniques, indicating increased sensitivity of the model to rare arrhythmias.
- Category V continued to score relatively low, which is attributed to the very limited number of available samples.
- These results indicate that imbalance correction effectively improves performance; however, extremely rare cases require additional data or more advanced models.

In general, the experimental findings highlight the following:

- Signal purification contributes to significantly improving the stability of the model and raising the overall accuracy.
- Data imbalance clearly negatively affects the detection of rare systemic disorders if not treated.
- The combination of data augmentation and category weighting achieves more balanced and appropriate performance for clinical applications.
- The proposed CNN-BiLSTM framework is performing competitively compared to recent studies, making it a strong foundation for the development of intelligent systems of heart rhythm disorders in realistic environments.

5.2. **Discussion.**

5.2.1. *The effect of pretreatment compared to previous work.* Experimental results showed that the integration of advanced pretreatment techniques, particularly noise removal using Daubechies-4, and normalization Z-Score, led to a noticeable improvement in the performance of the CNN-BiLSTM model, as the overall accuracy increased in the basic model to 88.53% after application of the purification. This improvement reflects the pivotal role of signal quality in enhancing the stability of the model and improving its generalization ability. These results are consistent with what is mentioned in the modern literature, where multiple studies such as Hassan et al. (2022) and Sattar et al. (2024) Noise and capacitive differences between patients are one of the most important factors that negatively affect the performance of deep learning models when analyzing ECG signals.

However, many of these studies rely on traditional filters or limited processing strategies, which restrict their effectiveness when dealing with realistic, high-noise signals. In contrast, the current work is characterized by a more integrated pre-treatment framework that maintains clinical morphological properties, in particular the QRS complex, while minimizing the distortion resulting from purification. This approach has significantly improved the taxonomic performance of common categories, which justifies the relative superiority of the proposed model compared to many previous works. Table 2 shows the comparison of the proposed model results with a number of previous studies related to the classification of cardiogenic disorders using electrocardiogram signals (ECG), in terms of the database used, the classification methodology, the category imbalance processing mechanism, and the precision of classification verified.

Although some previous studies have achieved higher accuracy rates compared to the proposed model, most of them relied on balanced data sets or ignored the problem of category imbalance, which could lead to an unrealistic performance assessment. In contrast, the proposed model in this study relied on the promotion of data and category weights

TABLE 2. Comparative analysis of the results of the classification of cardiac arrhythmias using ECG signals

Key Results	**Dataset**	**Handling of Class Imbalance**	**Model Used**	**Methodology**	**accuracy**	**Study**	**Scientific evaluation**
Enhanced accuracy over standalone CNN	MIT-BIH	Partial	CNN–BiLSTM	Hybrid spatial–temporal feature extraction	98.25%	Pandey et al. (2024)	An assessment based on relatively simplified data, with limited handling of imbalances [3]
Good performance on balanced subsets	MIT-BIH	No	CNN	Feature-based classification with imbalance analysis	99.4%	Ma et al. (2022)	High accuracy, but it doesn't reflect true performance due to ignoring imbalance [14]
Robust temporal modeling	MIT-BIH	Yes (SMOTE)	CNN–BiLSTM	End-to-end ECG classification	99.13%	Zhang et al. (2024)	Accuracy improved at the expense of clinical realism [19]
Proposed CNN–BiLSTM Model	MIT-BIH	Yes Augmentation + Class Weighting	CNN–BiLSTM	Wavelet denoising + Z-score normalization + hybrid spatial–temporal learning	88.53%		A realistic assessment that addresses genuine imbalances and enhances generalizability.

to address the real imbalance in the MIT-BIH database, which provides a more realistic and better generalization ability to generalize in practical applications.

5.2.2. *Addressing the imbalance of categories compared to traditional models.* The problem of the imbalance of categories is one of the most prominent challenges in the ECG databases. When comparing the results of this work with studies such as Ye et al. (2025) and Lamba et al. (2025), it turns out that many of the previous models either ignored the imbalance or treated it with only one technology, such as Oversampling or Class Weighting. On the other hand, this research adopted a double strategy that combines the weighting of categories and data, which allowed to improve the sensitivity of the model to the rare categories without the great sacrifice of total accuracy. The significant improvement in the retrieval value of the category A (Premature Atrial Beat) is a strong indicator of the effectiveness of this approach.

5.2.3. *The CNN-BiLSTM Hybrid Structure.* Many recent studies have relied on deep learning models, particularly convolutional and iterative networks, for classifying electrocardiogram (ECG) signals, such as Thota et al.'s A Lightweight CNN-Attention-BiLSTM Architecture and Sun et al.'s An Arrhythmia Classification Model Based on CNN-LSTM. Other works, such as Zhang et al. and Patnaik et al., have focused on hybrid and enhanced models using attentional or multi-input mechanisms. While these studies have achieved good accuracy, most have not explicitly addressed the problem of data imbalance, which limits the model's effectiveness in detecting rare conditions.

On the other hand, systematic review studies, such as Silva et al.'s A Systematic Review of ECG Arrhythmia Classification and Rahman et al., have also focused on this area. In A Systematic Survey of Data Augmentation for ECG Signals, recent research trends were analyzed, highlighting open challenges, primarily data imbalances, noise, and limited practical applicability of models. However, these studies remained within the theoretical and analytical framework, failing to provide practical implementation models or integrated systems that could be evaluated empirically.

In contrast, the current research offers a more comprehensive contribution by combining the findings of these separate studies. A hybrid model (CNN-BiLSTM) was adopted, capable of extracting the morphological and temporal characteristics of ECG signals, while implementing effective strategies to address data imbalances, such as artificial data augmentation and class weighting. Furthermore, this work goes beyond disconnected analysis, supporting practical applicability within smart healthcare systems. This makes it a direct and practical response to the research gaps identified by previous studies, enhancing its reliability and potential for adoption in medical IoT applications.

5.2.4. *Explanation of the performance of the very rare categories.* Despite the overall improvement, the Category V (early ventricular contraction, PVC) continued to show less performance compared to the rest of the categories. This can be explained by the very limited number of samples available for this category within the data set. Actually, PVC samples represent only 31 samples out of 40,937 heartbeats, equivalent to roughly 0.08% of the total data.

This severe imbalance of classes greatly affects the model's ability to learn effectively and generalize the patterns associated with this category. This finding suggests that improving the performance of such rare categories requires either the provision of additional data, the use of more advanced data generation technologies such as GANs, or the integration of multi-channel information, which represents promising directions for future research.

5.2.5. *Clinical and applied semantics.* The results show that the proposed model not only achieves high accuracy, but also a better balance between sensitivity and accuracy, which is critical in clinical systems. Natural. The model's relatively light structure, as well as training stability and convergence speed, makes the proposed framework suitable for integration into smart health control systems and the Internet of Things (IoT-Based Healthcare Systems) [2], [3].

The suggested CNN–BiLSTM model for ECG classification shows promise, although there are a few drawbacks to take into account. First, the study's dataset is extremely unbalanced, especially for minority classes like Premature Ventricular Contraction (PVC), which could have a detrimental effect on the model's capacity to generalize to all categories. Second, the model's application to other real-world clinical datasets with varied characteristics may be limited because it was only tested using the MIT-BIH Arrhythmia Database. Third, the system does not accurately represent real-time situations since it uses pre-segmented ECG signals based on annotated R-peaks. Lastly, the model continues to perform worse for uncommon classes, suggesting that future work will require larger datasets and more reliable data balancing strategies.

Limitations of the Proposed Method. Despite the promising performance of the proposed CNN–BiLSTM model in the ECG signaling, there are several limitations to consider. First, the data set used in this study suffers from a clear imbalance between groups, especially low-representation groups such as early ventricular contraction (PVC), which may affect the model's ability to effectively generalize all categories. Second, the model was evaluated using only one data set (MIT-BIH database for cardiac arrhythmia), which could limit the ability to generalize results to real clinical environments containing different levels of noise and patient variation. Third, the system relies on pre-cut and centered ECG signals around the pre-determined R tops, which does not fully reflect realistic scenarios that require automated partitioning. Finally, although the model is generally performed well, it still shows a decrease in accuracy for rare categories, indicating the need to use more advanced technologies to address data imbalances, as well as rely on larger and more diverse data sets in future businesses.

6. **Conclusion.** This research provided an integrated and effective framework for disorders based on ECG signals, by integrating advanced pretreatment techniques with a hybrid deep learning model that combines CNN and BILSTM. The standard MIT-BIH arrhythmia database database-based results showed that signal pretreatment—especially noise removal using the waveform and normalization with Z-Score—plays a pivotal role in improving the stability of the model and its ability to generalize. The study also showed that addressing the problem of category imbalance is a critical component of the model's sensitivity to rare system disorders, as the combination of synthetic data increase and category weighting has resulted in more balanced and clinical performance, compared to or those basic models.

Which depends on only one strategy? The proposed model achieved competitive overall accuracy where it reached 88.53%, with a noticeable improvement in retrieval metrics and the F1 scale for low-represented disease classes. These results confirm that the proposed framework is not only able to achieve high performance statistically, but also has the practical ability to be applied in smart monitoring and the Internet of Things, with promising potential for semi-interpretation and support for early warning systems. In light of this, this work can be considered an important step towards developing reliable and expandable diagnostic systems that contribute to improving the quality of heart health care and reducing the risks associated with delayed diagnosis of heart rhythm disorders.

REFERENCES

[1] S. S. Hussein, L. A. Tawfeeq, and S. S. Altyar, "Transforming Cardiac Diagnostics: Advanced Deep Learning Methods for Early Heart Disease Detection," *J. Inf. Hiding Multimed. Signal Process.*, vol. 16, no. 4, pp. 1085–1092, 2025.

[2] A. J. Prakash and M. Atef, "A lightweight deep learning approach for patient-specific electrocardiogram beat classification using local and long-term dependencies," *Eng. Appl. Artif. Intell.*, vol. 152, p. 110754, 2025, doi: 10.1016/j.engappai.2025.110754.

[3] V. Pandey et al., "Enhancing heart disease classification with M2MASC and CNN-BiLSTM integration for improved accuracy," *Sci. Rep.*, vol. 14, no. 1, p. 24221, 2024, doi: 10.1038/s41598-024-74993-2.

[4] M'hame et al., "Advanced Cardiac Monitoring via IoT: A CNN-Based Approach for Real-Time ECG Arrhythmia Classification," *Sensors*, 2024.

[5] Sun et al., "An Arrhythmia Classification Model Based on CNN and LSTM for IoT Healthcare Applications," *IEEE Access*, 2024.

[6] Thota et al., "A Lightweight CNN-Attention-BiLSTM Architecture for ECG Signal Classification," *Appl. Sci.*, 2024.

[7] Ye et al., "Hybrid CNN-BLSTM Architecture for Classification of Cardiac Arrhythmias," *Biomed. Signal Process. Control*, 2025.

[8] Makhir et al., "IOT Based ECG: Hybrid CNN-BILSTM Approach for Arrhythmia Detection," *Int. J. Adv. Comput. Sci. Appl.*, 2024.

[9] Do et al., "Data Augmentation for 12-Lead ECG Beat Classification," *Comput. Biol. Med.*, 2023.

[10] Hizem et al., “Reliable ECG Anomaly Detection on Edge Devices for IoT Healthcare Systems,” *IEEE Internet Things J.*, 2023.
[11] Liu et al., “Automatic Classification of Arrhythmias Using Deep Neural Networks,” *Front. Physiol.*, 2023.
[12] Rincon et al., “An IoT and Fog Computing-Based Monitoring System for Cardiac Arrhythmia Detection,” *Sensors*, 2022.
[13] Sattar et al., “Cardiac Arrhythmia Classification Using Advanced Deep Learning Techniques,” *Healthc. Inform. Res.*, 2024.
[14] Ma et al., “Deep Learning-Based Data Augmentation and Model Fusion for ECG Arrhythmia Classification,” *Comput. Biol. Med.*, vol. 145, p. 105500, 2022.
[15] Rahman et al., “A Systematic Survey of Data Augmentation for ECG Signals,” *IEEE Access*, 2022.
[16] Patnaik et al., “Attention Enhanced Hybrid Deep Learning Model for ECG Arrhythmia Classification,” *Biomed. Signal Process. Control*, 2023.
[17] Panwar et al., “Integrated Portable ECG Monitoring System with Real-Time Arrhythmia Detection,” *J. Med. Syst.*, 2023.
[18] Gradl et al., “Real-Time ECG Monitoring and Arrhythmia Detection Using Android Smartphones,” *Procedia Comput. Sci.*, 2012.
[19] Zhang et al., “Arrhythmia Classification Based on Multi-Input Neural Networks,” *Expert Syst. Appl.*, 2024.
[20] Farag, “A Tiny Matched Filter-Based CNN for Inter-Patient ECG Classification,” *Biomed. Signal Process. Control*, 2021.
[21] Wang et al., “Atrial Fibrillation Detection and ECG Classification Based on Deep Neural Networks,” *IEEE J. Biomed. Health Inform.*, 2022.
[22] Hassan et al., “Classification of Cardiac Arrhythmia Using a Convolutional Neural Network,” *Diagnostics*, vol. 12, no. 4, p. 951, 2022.
[23] Silva et al., “A Systematic Review of ECG Arrhythmia Classification Methods,” *Artif. Intell. Med.*, 2022.
[24] Lamba et al., “FADLEC: Feature Extraction and Arrhythmia Detection Using Deep Learning for ECG Classification,” *Biomed. Signal Process. Control*, 2025.
[25] M. Hamad, “Remote ECG Signal Monitoring and Classification Using Machine Learning,” *J. Healthc. Eng.*, 2022.