Regular Article

Electrocardiogram Based Heartbeat Detection Using Deep Learning

Minh Tuan Nguyen, Anh Nguyen

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Correspondence: Minh Tuan Nguyen, nmtuan@ptit.edu.vn
Communication: received 17 March 2025, revised 23 April 2025, accepted 27 May 2025
Online publication: 01 June 2025, Digital Object Identifier: 10.21553/rev-jec.401

Abstract– **Cardiovascular diseases remain a leading cause of death worldwide, which results in important requirements for early and accurate arrhythmia diagnosis. This work proposes a novel design of automated heartbeat detection, which consists of a convolutional neural network (CNN) and three-channel images using the electrocardiogram (ECG) signals. A combination of various preprocessing is applied for the elimination of interferences of the ECG signals such as band-pass filtering and wavelet transform for R-peak identification using a sliding window. Multimodal image fusion method is used to construct three-channel images from different grayscale images, which are transformed from the heartbeats by three transformation techniques namely Gramian angular field, Markov transition field, and Recurrence plot. Grid-search based optimization method in combination with 5-fold cross validation procedure are implemented for selection of the optimal hyper-parameters of the CNN models using the input three-channel images. The proposed algorithm including CNN models and markov transition field images is estimated the detection performance using 5-fold cross validation, which produces average accuracy of 99.63%, precision of 99.41%, recall of 99.52%, and F1-score of 99.64%. The relatively high performance of the proposed algorithm confirms the effectiveness for the arrhythmia recognition on the ECG signals.**

Keywords– **Cardiovascular diseases, electrocardiogram noise filter, multimodal image fusion, deep learning, heartbeat classification.**

1 Introduction

Cardiovascular diseases (CVDs) are among the leading causes of death worldwide, encompassing conditions such as coronary artery disease, heart failure, and arrhythmias. There are a large number of factors which have important impact on these heart diseases such as hypertension, diabetes, obesity, and an unhealthy lifestyle [1]. Early detection and proper management of CVDs are essential to reduce mortality rates and improve patient quality of life. Nowadays, advances in artificial intelligence and signal processing applied for medical fields, have significantly enhanced the diagnosis and treatment of the above diseases.

Electrocardiogram (ECG) has been widely used for non-invasive techniques to monitor heart activities [2]. Indeed, ECG provides valuable information about heart rhythms, conduction abnormalities, and possible cardiovascular conditions such as arrhythmias and myocardial infarction. Due to its simplicity and effectiveness, ECG is commonly used in clinical settings and wearable health-monitoring devices, which adopt state-of-the-art deep learning and machine learning to enable automated ECG analysis, improving accuracy and efficiency in detection of the heart diseases [3].

Despite significant technological advances, the utility of ECG signals still presents several challenges. One major issue is the susceptibility to various sources of noises, including external interferences such as baseline drift caused by medical equipment inaccuracies and internal disruptions from physiological activities like muscle contractions. These noise artifacts can obscure critical details, such as the exact position of ECG peaks, making arrhythmia detection more complex and potentially leading to misinterpretations during clinical assessments. Another challenge lies in the intricate nature of ECG signals, which require specialized expertise to analyze correctly. This complexity makes it difficult for non-experts to leverage ECG data effectively to detect heart conditions. Consequently, the demand for automated solutions has grown substantially. Over time, numerous noise reduction techniques have been introduced such as conventional filtering methods, adaptive filtering, and singular value decomposition-based approaches [4]. Clearly, these techniques have demonstrated their effectiveness but inherent drawbacks are still available. Indeed, standard filters such as low-, high-pass, and notch filters may inadvertently eliminate important signal components during noise reduction process, which definitely results in signal distortion [5]. Although adaptive filtering is powerful, it can be computationally intensive and highly dependent on precise parameter adjustments.

In recent years, machine learning has become increasingly important in arrhythmia diagnosis, owing to its ability to detect complex patterns in large ECG datasets. There are numerous studies, which prove the effectiveness of machine learning methods for classification of various arrhythmia types. Decision trees, support vector machines (SVM), and k-nearest neighbors (KNN) have been successfully used to identify abnormal heart rhythms with promising results. The authors of [6] propose an ensemble-based SVM model to categorize heartbeats into four classes using features such as wavelets, high-order statistics, R-R intervals, and morphological characteristics. This approach outperforms the other baseline models such as individual SVM, Random Forest, and KNN, achieving the highest accuracy of 94.4%. In [7], Fourier decomposition and phase transform are applied to extract key features from ECG signals, which capture both inter- and intra-beat variations. The proposed method produces an impressive accuracy of 97.92% on the MIT-BIH dataset. However, there are significant challenges which need to be solved using machine learning models and complex biomedical data such as ECG signals. One of main limitations is their dependence on extensive feature engineering required human expertise to manually extract relevant features. Moreover, feature engineering is a time-consuming process and prone to errors. In addition, traditional machine learning approaches struggle to capture the inherently temporal dependencies of the ECG signals, hindering their ability to generalize across various arrhythmia types.

To overcome the limitations of traditional machine learning methods, deep learning has emerged as a superior approach for the arrhythmia detection. Unlike conventional machine learning techniques, deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM) can automatically learn features directly from raw ECG signals, eliminating the need for manual feature extraction. In [8], discrete wavelet transform is used to extract statistical features, followed by classification using a multilayer perceptron neural network with backpropagation. This approach effectively distinguishes between normal and abnormal ECG signals from the MIT-BIH Arrhythmia and Apnea ECG databases, which obtains an overall accuracy of 94.44%. The authors of [9] demonstrate the effectiveness of a 1D-CNN model including three convolutional layers, max-pooling layers, and dense layers to extract nonlinear features from ECG data and classify them into five categories. The evaluated accuracy of this study is 97.36% on the MIT-BIH database with 47 subjects using 5-fold cross-validation method. However, deep learning models show important drawbacks when using noisy ECG signals or incomplete data. More specific, the utility of noisy ECG signals as the input of the deep learning model possibly results in low recognition performance in practical scenarios. Besides, it is necessary to perform extensive computations for the deep learning model using directly the raw ECG signals, which leads to long training time and high computational cost. Consequently, this poses an essential obstacle for the real medical applications, which requires frequently rapid process with limited computational resources.

In this work, we propose a novel algorithm for the arrhythmia recognition including three-channel images as the input of the deep learning models. The images are a combination of three transformed grayscale images by Gramian angular field (GAF), Markov transition field (MTF), and Recurrence plot (RP) methods to improve the feature representation in terms of heartbeat types in the ECG signals. Mover, heartbeats are selected carefully by a sliding window of 360 samples from the ECG signals which are preprocessed by band-pass filtering and wavelet transform to effectively remove interferences. The main contributions of our work are as follows:

- Development of an efficient noise reduction by utility of combined filtering methods namely band-pass filtering and Wavelet transform to improve the quality of ECG signals while maintaining critical features of QRS complexes, P, and T waves.
- Improvement of the image feature representation by the utility of the multimodal image fusion method to integrate different gray images as the outputs of various transformation techniques such as GAF, MTF, and RP.
- Proposal of a high-performance arrhythmia recognition algorithm using state-of-the-art techniques and ECG signals, which is reliable for clinical application in healthcare environments.

The rest of the paper is organized as follows: The description of the data used in this study is given in Section II followed by the methodology proposed in Section III. The simulation results and discussion are presented in Sections IV and V. Section VI shows the conclusive remarks of this work.

2 Data

Following the guidelines set by the Association for the Advancement of Medical Instrumentation (AAMI), we utilize the MIT-BIH Arrhythmia dataset from PhysioNet databases for this work [10]. The dataset is widely used as benchmark data for classification and validation of heartbeat recognition ensuring the reliability and comparability of different approaches. There are 48 ECG recordings from 47 subjects, each lasting 30 minutes and sampled at 360 Hz on two leads (V and II) in the MIT-BIH database. In this study, we specifically consider lead II due to clear representation of key ECG waveform components, which makes it fit in ECG beat extraction and classification. In addition, the dataset includes detailed annotations of arrhythmia types and R-peak locations, which are meticulously reviewed and confirmed by reliable cardiologists. These high-quality annotations are essential to ensure precise labeling, facilitate robust model training, and enhance the accuracy of arrhythmia detection [3–6].

3 Research Method

The proposed methodology is shown in Figure. 1, which consists of three main phases: data and preprocessing, transformation, and model evaluation.

- In the data and preprocessing phase, the raw ECG signals are processed by a series of enhancement techniques to obtain high-quality signals for subsequent

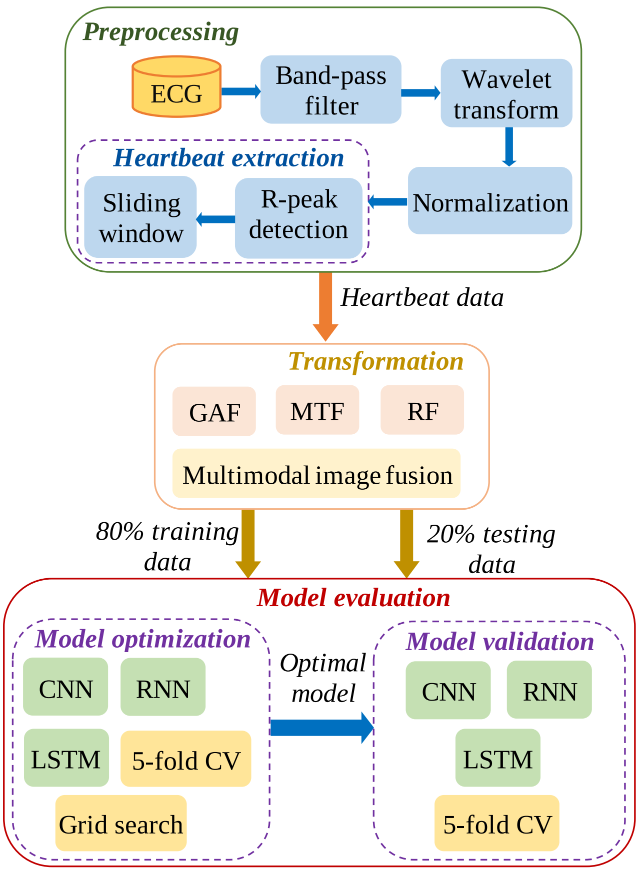


Figure 1. Method diagram

analysis. At first, bandpass filtering is applied to remove baseline wander and high-frequency noise. Then, the wavelet transform is employed for denoising the signal while preserving key morphological features. Furthermore, Min-Max normalization is used to scale the signal values between 0 and 1, ensuring consistency across the dataset. The R peak detection and optimized segmentation window are performed to identify critical points of the ECG signal and to extract individual heartbeats, respectively, generating a structured and standardized dataset.

- In the transformation phase, heartbeats are converted into different visual representations to improve spatial and temporal dependencies. Indeed, three transformation techniques namely GAF, MTF, and RP are employed to provide a richer feature representations, allowing deep learning models to better capture patterns and relationships within the ECG signals. To further enhance feature extraction, a multimodal image fusion (MIF) strategy is used to integrate these representations into a unified dataset, ensuring a comprehensive input for classification.

- In model evaluation stage, deep learning models such as CNN, LSTM, and recurrent neural network (RNN) are fine-tuned using grid search and 5-fold cross-validation to determine the best hyperparameters. The optimal models are then validated their performance in terms of heartbeat classification on the testing data.

3.1 Preprocesing and heartbeat extraction

The descriptions of various techniques processed for the extraction of heartbeats are as follows:

3.1.1 Preprocessing: A comprehensive denoising method for ECG signals, which includes two advanced filtering techniques such as band-pass filtering and wavelet transform, is adopted in this work [11]. These techniques are carefully selected to target specific types of noise while maintaining vital ECG features, which results in an improvement of the subsequent analysis performance. Initially, band-pass filtering is applied to eliminate both low-frequency baseline wander and high-frequency noise within a frequency range of [0.5-40] Hz. Then, the Wavelet transform is applied to further refine the signal quality, which generates the cleaned ECG signals for further implementation of the Min-Max normalization to scale the ECG signals to a range of [0-1].

3.1.2 Heartbeat extraction: The Pan-Tompkins algorithm has been used in many studies to detect QRS. However, it is not effective in handling noisy or low-quality ECG signals. As demonstrated in [12], the performance of the Pan-Tompkins algorithm significantly degrades under such conditions. In contrast, X-wave QRS detection (XQRS) is suitable for the ECG signals including different challenges such as noise, baseline wander, and QRS morphologies, which makes it become a highly reliable tool for R-peak detection in various datasets and scenarios. Therefore, in this research, we use XQRS method to detect the positions of the R-peaks in ECG signals [13]. Indeed, the R-peak, which represents the highest point in a QRS complex of an ECG waveform, corresponds to the depolarization of the ventricles. Different techniques of signal processing and mathematical methods are employed for the XQRS to discern key features of the QRS complex, such as amplitude, duration, and shape.

Heartbeats are then extracted by a sliding window with a dimension of 360 samples using the above identified R-peaks [14]. The rational behind the use of the above window is that its size corresponds to a 1-second window at a sampling rate of 360 Hz, which closely approximates the average duration of a single heartbeat. Firstly, the ECG signal is divided into multiple segments, in which each segment has an R-peak as a center to capture the entire vital features of the heartbeats. Then, a window size of 360 samples around an R-peak from P-179 to P+180 is applied to extract heartbeats, which minimizes the risk of incomplete waveform coverage and helps the model learn the full heart rate morphology. As a result, critical regions for the identification of the cardiac arrhythmias, including the QRS complexes, P-waves, and T-waves are definitely included in this window size. Here, we select a conventional 360-sliding window due to its effectiveness in terms of heartbeat extraction and improvement of the final arrhythmia detection performance. Besides, we adopt the synthetic minority oversampling technique (SMOTE) to generate additional samples to suppress imbalanced data problem. This technique addresses class imbalance by the generation of synthetic samples for minority classes which results in a more balanced distribution of the outcomes. The synthetic samples are produced by interpolating between existing minority class instances. A minority class instance is selected with its nearest neighbors, which creates the synthetic samples along the line segments connecting the instance to its neighbors. Consequently, the minority class representation is significantly increased without merely duplicating existing samples, which leads to mitigation of class imbalance and improvement of the model performance.

3.2 Transformation

We employ three different transformation methods to convert raw ECG signals into images: GAF, RP, and MTF to improve the quality of the transformed images. Then, the MIF technique is implemented for the combination of the above transformation outputs to generate informative heartbeat-based images, which preserve both statistical and temporal features. The combined images are further fed into the deep learning models to estimate the performance of those feature images and intelligent algorithms.

1) *GAF Transformation:* The heartbeats are encoded to the angular coordinate system by the GAF. Firstly, the heartbeats are normalized from 0 to 1, followed by mapping them to the angular transformation in which the time and amplitude values define the radial and angular components, respectively. The GAF ensures the maintenance of the spatial relationships among heartbeat samples and preserves the temporal dependencies in the heartbeat, which allow to better visualization of the heartbeat patterns [15].
2) *RP Transformation:* This method converts the heartbeats into images by capturing their recurrence behavior over time. Due to non-stationary characteristic of the heartbeats, the RP improves significantly the visualization of how similar heartbeat patterns repeat at different time intervals. Each point in an RP image represents the similarity between two time instances in the heartbeat, which identify periodic structures and anomalies. The RP transformation is particularly useful for arrhythmia detection and other irregularities in the heartbeat patterns due to its ability of emphasis on the repetitive features within the heartbeats [16].
3) *MTF Transformation:* First-order Markov chains based probabilistic approach, which includes partition of a heartbeat into quantile-based bins and construction of a weighted adjacency matrix representing transition probabilities between different signal states, is employed to encode the heartbeats. Consequently, Markov transition field matrix captures spatial dependencies of a heartbeat while preserving the temporal dynamics. Moreover, the transition probabilities across the spatial domain is extended by the MTF to maintain the essential heartbeat characteristics for the classification [16].

3.3 Model evaluation

There are three deep learning models namely CNN [17], RNN [18], and LSTM [19] considered as the intelligent algorithms for the classification of various heartbeats. To improve the model performance, we implement the gird search-based optimization method to obtain the optimal hyperparameters of such models and then validate the selected models as follows:

3.3.1 Model Optimization: It is clear that optimization plays an important role of the model performance improvement and overfitting avoidance [20, 21]. Indeed, key parameters of the CNN model such as the learning rate, batch size, optimizer, and number of epochs are necessary to fine-tuned for the maximization of the feature extraction efficiency. Furthermore, RNN and LSTM, which are designed for sequential data processing, require additional tuning for the number of units, dropout rate, and sequence length to effectively capture temporal dependencies in the heartbeats. In this study, we apply systematic optimization process including grid search and 5-fold cross-validation procedure to address the optimal parameter of these models on the training data. Consequently, the optimal models, which are able to learn robust representations of the heartbeat waveforms, result in better accurate arrhythmia detection.

Here, we select different range of values for the optimization of model parameters and structures. There are optimizers of [Adam, SGD, RMSprop], batch sizes of [50, 75, 100], epochs of [50, 60, 70, 80, 90, 100], learning rate of [0.0001, 0.0002, 0.0003, 0.0004, 0.0005]. Moreover, we define 5 blocks including a convolutional layer, a ReLU, and a Maxpooling layer. Consequently, 5 CNN structures, which consist of the number of blocks ranging from 1 to 5, is considered for the identification of the optimal CNN model by the grid search method. Similarly, we investigate 5 structures of LSTM models and 5 RNN structures in which each model structure contains a number of LSTM or Recurrent layer ranging from 1 to 5 for the selection of the optimal LSTM and RNN models. A total of 15 structures with 270 combinations of parameter values result in 4050 models of CNN, LSTM, and RNN, which are then put into the grid search algorithm.

3.3.2 Model Validation: The validation-based statistic maner is applied for estimation of the optimal deep learning models on the testing data, which make the proposed algorithm become reliable application in the practical environment. Here, the testing heartbeat data is partitioned into five subsets, in which four subsets are used for training and remaining subset is for testing in each iteration. This process is repeated five times so that all individual subsets serves as the testing subsets. Besides, the mean performance metrics are then computed for the comparison with the existing methods.

4 Experiment Results

4.1 Performance measurement

We use accuracy, recall, precision, and F1-score for model performance estimation in this work. Accuracy measures the proportion of heartbeats identified correctly. Recall represents the proportion of heartbeats that are identified correctly while precision shows the fraction of the model's heartbeat classifications that are correctly addressed. The performance metrics are given as follows for individual types of heartbeats. The final performance is then computed as the mean of all heartbeat types:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (1)$$

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

$$Precision = \frac{TP}{FP + TP}, \quad (3)$$

$$F1\_score = \frac{2 * Recall * Precision}{2 * (Recall + Precision)}, \quad (4)$$

where TP, FN, TN, and FP are true positive, false negative, true negative, and false positive values.

4.2 Preprocessing and heartbeat extraction

4.2.1 Preprocessing: Figure 2 show an example of the preprocessed ECG signals, known as the outputs of different preprocessed techniques in subfigures such as raw ECG data in subfigure a), output signal of the band-pass filter in subfigure b), reconstructed signal by the wavelet transform in subfigure c), and the normalized signal in subfigure d). Obviously, the signals for the extraction of high-quality heartbeats are significantly improved by the above preprocessed techniques.

4.2.2 Heartbeat segmentation: The total number of heartbeats extracted from the 48 records is as follows: 72.471 for Normal (N), 2.223 for Supraventricular (S), 5.788 for Ventricular (V), 641 for Fusion (F), and 6.431 for Unknown (Q), respectively. SMOTE is then applied to address imbalanced data, which results in N of 72.471, S of 30.000, V of 20.000, F of 20.000, and Q of 10.000. The dataset of five heartbeat types is then separated into training data of 80% and testing data of 20% for further steps.

4.3 Transformation

The MIF generates the images which consist of three orthogonal channels corresponding to GAF, RP, and MTF images used as the input of various deep learning algorithms. Clearly, three-channel image is a composite representation, which includes the grayscale images converted by three transformation techniques using the heartbeats processed by different method such as wavelet transform and band-pass filtering. Table I shows the transformed images of different heartbeat types.

4.4 Model evaluation

4.4.1 Model optimization: Table II and Table III present the optimal structures and parameters of the deep learning models, which are then validated their performance on the testing data using 5-fold cross validation procedure.

4.4.2 Model validation: The mean performance comparison of different deep learning models such as LSTM, RNN, and CNN is given in Table IV. The highest average performance is obtained by the CNN model with an accuracy of 99.62%, precision of 99.41%, recall of 99.52%, and F1-score of 99.64%, which is selected as the proposed algorithm for the heartbeat detection.

5 Discussions

A combination of preprocessing techniques including band-pass filtering, wavelet transform, and normalization are used in this work. These techniques are carefully selected to target specific types of noise while maintaining vital ECG features, which results in the improvement of the subsequent analysis performance. Clearly, slow drifts and high-frequency artifacts, which distort important signal components, are efficiently removed to obtain the clean ECG signal without unwanted noise. Indeed, Figure 2b shows the output signal of a band-pass filter with a cut-off frequency of [0.5-40] Hz, which significantly reduces unwanted frequency components, although some residual noise remains. The filtered signals are further decomposed into multiple frequency bands by the wavelet transform to reduces remaining noise while preserving essential features such as QRS complexes, P waves, and T waves. Clearly, the wavelet approach is especially effective in removing non-stationary noise that may not have been fully addressed by the band-pass filter. Figure 2c proves the effectiveness of the wavelet transform for the signal denoising. Here, the filtered signal is enhanced by smoothing out residual noises while preserving essential features. The normalization is mandatory procedure to standardize the data and improve the performance of different deep learning models. It is obvious that we we are able to minimize the influence of varying amplitudes and enhance the model's ability to identify patterns and make accurate predictions by the use of normalization process. Consequently, This the ECG data is uniformly scaled and ready for further analysis, classification tasks.

There are three transformation techniques employed in this work to improve the spatial and temporal characteristics of the heartbeats, which make them better to use as the input of deep learning models. Moreover, each method preserves distinct statistical properties of the ECG data, contributing to a more comprehensive feature representation and lossless conversion. Then, MIF is adopted to combine three gray images generated by the above transformation methods to form a three-channel image used as three orthogonal channels, similar to the three colors in the RGB (red, green, and blue) image space. However, unlike conventional RGB

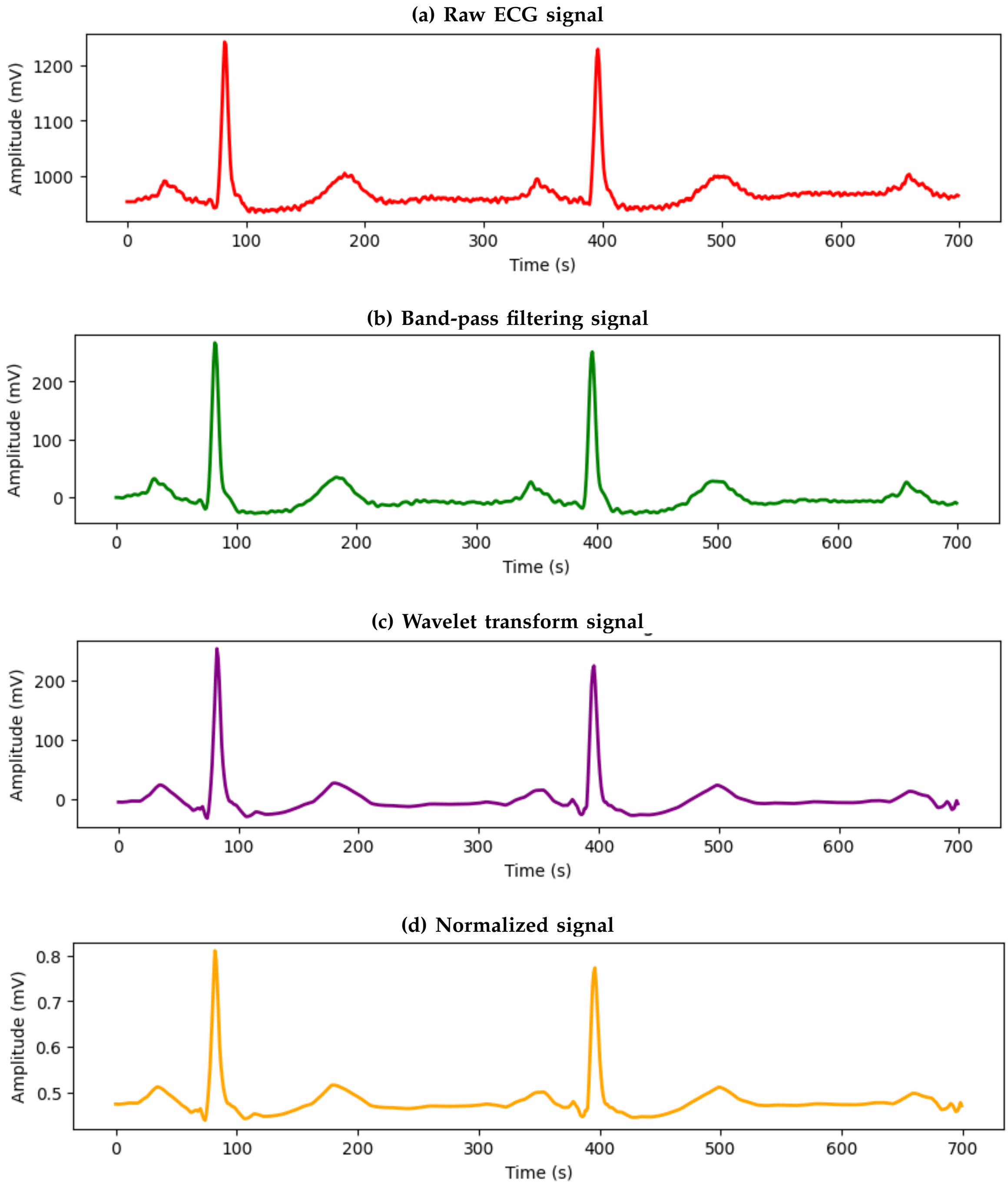


Figure 2. Examples of preprocessed ECG signal. (a) Raw ECG signal; (b) Band-pass filtering signal; (c) Wavelet transform signal; (d) Normalized signal.

image conversion, MIF algorithm generates all three grayscale images directly from the heartbeat data using different statistical methods. As a result, the three-channel images capture better statistical dynamics of the heartbeats, which provide more comprehensive data representations. As a result, three-channel images correspond to different heartbeats, which are utilized as the input of various deep learning models for the arrhythmia diagnosis.

We implement CNN, RNN, and LSTM models for comparison with existing methods and proposal of the optimal heartbeat detection algorithms. CNN is suitable for automatically extracting spatial features, particularly from key waveform components namel QRS complexes, P-waves, and T-waves. In addition, good ability of the CNN to recognize intricate spatial relationships improves significantly arrhythmia detection. Another model known as RNN is well-suited for capturing temporal dependencies in ECG waveforms, enabling a deeper understanding of sequential heartbeat patterns. LSTMs, which is designed to retain long-term dependencies, further enhance temporal modeling in heartbeat analysis. The memory mechanisms of LSTM allow for better handling of extended sequences, improving classification accuracy and reliability. It is noteworthy that hyper-parameters such as batch size, epochs, learning rate, and optimizer are critical to ensure robust heartbeat interpretation and improve the

Table I
Transformed Images of Five Heartbeat Types

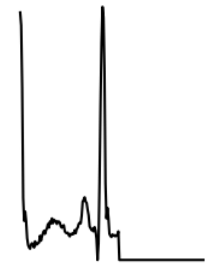
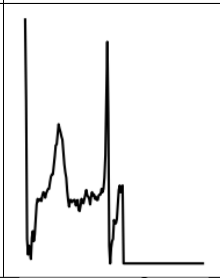
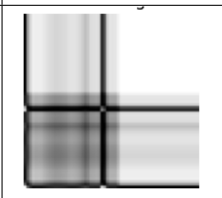
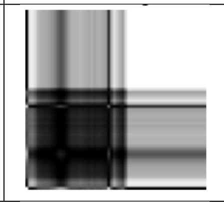
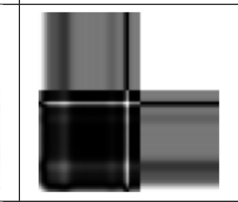
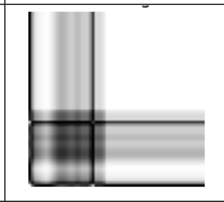
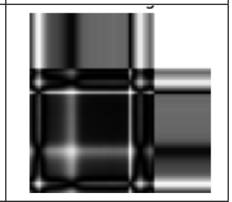
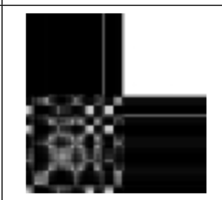
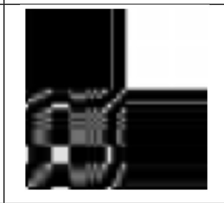
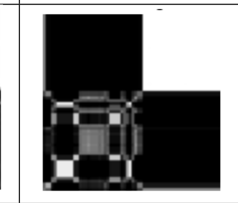
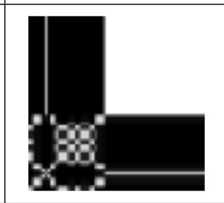
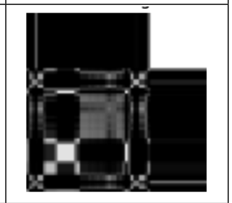
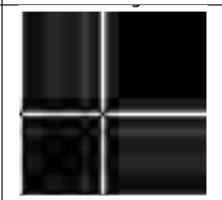
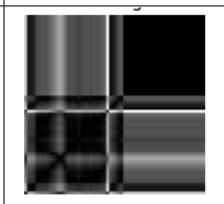
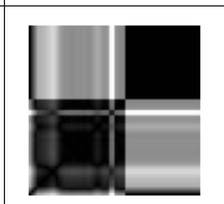
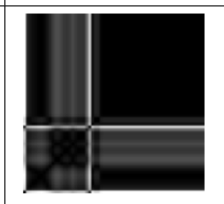
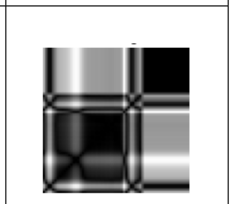
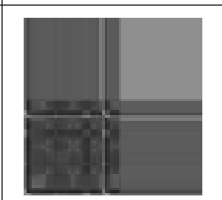
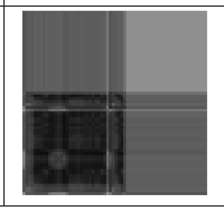
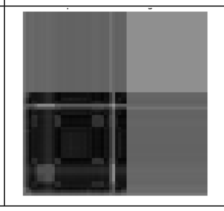
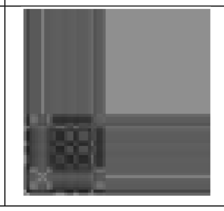
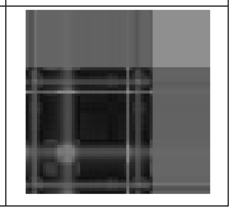
Categories	N	F	V	F	Q
Heartbeats					
GAF Images					
RP Images					
MTF Images					
Triple Chanel Image					

Table II
The Structure of Deep Learning Models

Model	Layer	Number
CNN	Convolutional layer	3
	ReLU	3
	Max pooling	3
	Dropout	1
	Fully connected	2
	Softmax	1
LSTM	LSTM layer	3
	Dropout	1
	Fully connected	2
	Softmax	1
RNN	Recurrent layer	3
	Dropout	1
	Fully connected	2
	Softmax	1

Table III
The Optimal Parameters of the Selected Models

Model	Parameter	Value
CNN	Optimizer	Adam
	Batch Size	100
	Epochs	50
	Learning rate	0.0001
LSTM	Optimizer	Adam
	Batch size	100
	Epochs	70
	Learning rate	0.0003
RNN	Optimizer	Adam
	Batch size	75
	Epochs	60
	Learning rate	0.0002

final performance of the proposed algorithm. Hence, the grid search combined with 5-fold cross validation procedure is implemented to address the optimal set of hyper-parameters for these deep learning models.

Table IV shows the average performance of three deep learning models on the testing data. Obviously,

Table IV
Performance of the Deep Learning Models on the Testing Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	97.32	97.39	98.39	97.32
RNN	98.43	98.42	98.36	98.44
CNN	**99.62**	**99.41**	**99.52**	**99.64**

Table V
Mean Accuracy Comparison of Proposed Algorithm with Existing Works

Model	Mean accuracy	Mean precision	Mean recall	Mean F1-score
[22]	97.78	71.3	60.98	32.87
[23]	98.35	98.41	98.30	98.36
Our	***99.62***	***99.41***	***99.52***	***99.64***

all three models produce relatively high performance in terms of heartbeat recognition with mean accuracy over 97%. Among these models, the highest performance is shown by the CNN model with average accuracy of 99.62%, precision of 99.41%, recall of 99.52%, and F1-score of 99.64%, which confirms the most effective approach of CNN, with respect to feature extraction and classification of the heartbeats. Hence, we select the proposed algorithm using CNN and a combination of transformed images using heartbeats extracted from the ECG signals.

Table V compares the average accuracy of the proposed method to that of the existing studies. Specifically, the proposed method achieves a remarkable average accuracy of 99.62%, which outperforms the reported accuracies of 97.78% and 98.35% in [22] and [23]. Therefore, the proposed algorithm is potential for practical use in the early detection and management of cardiovascular diseases.

6 Conclusions

CVD is one of the most dangerous heart diseases, which is the main cause of global mortality. Raid and correct diagnosis of CVDs play a vital role in healthcare systems, which provides essentially clinical decision-making for the experts and technicians in the practical hospital environments. Hence, the performance improvement of the arrhythmia detection is paid intensive attention from the medical researchers due to high classification performance resulting in avoidance of numerously unexpected deaths.

In this work, we proposed an effective algorithm, which is potential application for the clinic environments using deep learning. The proposed algorithm was designed with a CNN model and combined images using MIF method. Transformation techniques such as MTF, GAF, and RP are employed to convert the heartbeats into the grayscale images, which are then combined as three-channel images for the input of CNN model. Obviously, distinct statistical properties of heartbeats are captured significantly by three transformation techniques, which result in better feature representation and lossless conversion of the three-channel images constructed by MIF method. Moreover, essential characteristics of QRS complexes, T, and P waves are also maintained successfully by the combination of different preprocessing techniques namely band-pass filtering and wavelet transform. The validated classification performance with Ac of 99.62%, precision of 99.41%, recall of 99.52%, and F1-score of 99.64% on the testing data implies the effectiveness of our proposed algorithm to apply for the healthcare system, which provides essential support and reliable solution for the clinic experts with respect to early detection, proper treatment, and management of cardiovascular diseases in practical hospital environments.

The limitation of this work is the implementation of transformation technique to convert the ECG signals into multiple image formats, which increases the computational complexity. Moreover, the analysis is limited to a single dataset which raises concerns about the generalization of the results to other clinical scenarios.

Acknowledgment

This work was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant number 102.01-2023.03

References

[1] W. H. Organization, "Cardiovascular Diseases (CVDs)," Geneva, Switzerland, 2019, https://www.who.int/en/newsroom/factsheets/detail/cardiovascular-diseases-(cvds).

[2] A. Ameen, I. E. Fattoh, T. Abd El-Hafeez, and K. Ahmed, "Advances in ECG and PCG-based cardiovascular disease classification: a review of deep learning and machine learning methods," *Journal of Big Data*, vol. 11, no. 1, p. 159, 2024.

[3] N. Vinay, K. Vidyasagar, S. Rohith, P. Dayananda, S. Supreeth, and S. Bharathi, "An RNN-Bi LSTM based multi decision GAN approach for the recognition of cardiovascular disease (CVD) from heart beat sound: a feature optimization process," *IEEE Access*, vol. 12, pp. 65 482–65 502, 2024.

[4] O. Selma and A. Imteyaz, "Comparative Performance Analysis of Filtering Methods for Removing Baseline Wander Noise from an ECG Signal," *Fluctuation and Noise Letters*, vol. 23, no. 4, p. 2450046, 2024.

[5] T. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Translation Journal on Magnetics in Japan*, vol. 2, no. 8, pp. 740–741, 1987.

[6] S. K. Pandey, R. R. Janghel, and V. Vani, "Patient specific machine learning models for ECG signal classification," *Procedia Computer Science*, vol. 167, pp. 2181–2190, 2020.

[7] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, "Biometric identification from ECG signals using Fourier decomposition and machine learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.

[8] S. Mondal and P. Choudhary, "Detection of normal and abnormal ECG signal using ANN," in *Proceedings of the Advances in Intelligent Informatics, Smart Technology and Natural Language Processing: Selected Revised Papers*

from the Joint International Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2017). Springer, 2019, pp. 25–37.

[9] M. Chourasia, A. Thakur, S. Gupta, and A. Singh, "ECG heartbeat classification using CNN," in *Proceedings of the 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, 2020, pp. 1–6.

[10] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[11] D. Kim, K. R. Lee, D. S. Lim, K. H. Lee, J. S. Lee, D.-Y. Kim, and C.-B. Sohn, "A novel hybrid CNN-transformer model for arrhythmia detection without R-peak identification using stockwell transform," *Scientific Reports*, vol. 15, no. 7817, pp. 1–11, 2025.

[12] M. N. Imtiaz and N. Khan, "Pan-Tompkins++: A robust approach to detect R-peaks in ECG signals," in *Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 2905–2912.

[13] S. Dhyani, A. Kumar, S. Choudhury, C. Verma, and Z. Illés, "Assessing ECG-QRS signal detection algorithm chip and simulation on several FPGAs," *Discover Computing*, vol. 28, no. 1, pp. 1–12, 2025.

[14] O. A. Cárdenas, L. M. F. Nava, F. G. Castañeda, and J. A. M. Cadenas, "ECG Arrhythmia Classification for Comparing Pre-Trained Deep Learning Models," in *Proceedings of the 2022 19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, 2022, pp. 1–5.

[15] C. Lin, H. Lu, P. Sang, and C. Pan, "A knowledge embedded multimodal pseudo-siamese model for atrial fibrillation detection," *Scientific Reports*, vol. 15, no. 1, pp. 1–15, 2025.

[16] C. Wang, J. Ma, G. Wei, and X. Sun, "Analysis of Cardiac Arrhythmias Based on ResNet-ICBAM-2DCNN Dual-Channel Feature Fusion," *Sensors*, vol. 25, no. 3, pp. 1–21, 2025.

[17] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021.

[18] L. R. Medsker and L. Jain, *Recurrent neural networks: Design and Applications*. United States: CRC Press, Inc., 1999.

[19] A. Graves and A. Graves, "Long Short-Term Memory," *Supervised Sequence Labelling with Recurrent Neural Networks*, pp. 37–45, 2012.

[20] V. Satheeswaran, G. N. Chandrika, A. Mitra, R. Chowdhury, P. Kumar, and E. Glory, "Deep Learning based classification of ECG signals using RNN and LSTM Mechanism," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 4, pp. 332–342, 2024.

[21] P. N. Singh and R. P. Mahapatra, "A novel deep learning approach for arrhythmia prediction on ECG classification using recurrent CNN with GWO," *International Journal of Information Technology*, vol. 16, no. 1, pp. 577–585, 2024.

[22] L.-H. Wang, Y.-T. Yu, W. Liu, L. Xu, C.-X. Xie, T. Yang, I.-C. Kuo, X.-K. Wang, J. Gao, P.-C. Huang *et al.*, "Three-heartbeat multilead ECG recognition method for arrhythmia classification," *IEEE Access*, vol. 10, pp. 44 046–44 061, 2022.

[23] J. Cui, L. Wang, X. He, V. H. C. De Albuquerque, S. A. AlQahtani, and M. M. Hassan, "Deep learning-based multidimensional feature fusion for classification of ECG arrhythmia," *Neural Computing and Applications*, pp. 1–15, 2023.

Minh Tuan Nguyen received the B.S. degree from the Post and Telecommunications Institute of Technology, Hanoi, Vietnam, in 2004, the M.S. degree from Hanoi University of Science and Technology, Hanoi, Vietnam, in 2008, both in electronics and telecommunications engineering, and the Ph.D. degree at the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2018. He is with Posts and Telecommunications Institute of Technology. His research interests include network security, internet of things, biomedical signal processing, gene analysis, sentiment analysis, brain computer interface, machine learning, deep learning, optimization, and biomedical application design.

Anh Nguyen received a B.S. degree in computer science from the VNU University of Engineering and Technology – VNU-UET in 2013, the Ph.D. degree from the Energy Institute, Nanyang Technological University in 2018. He is with the Posts and Telecommunications Institute of Technology in Hanoi, Vietnam. His research interests are artificial intelligence applications, optimization.