HHS Public Access

Author manuscript

Intell Syst Appl. Author manuscript; available in PMC 2024 August 28.

Published in final edited form as:

Intell Syst Appl. 2024 June ; 22: . doi:10.1016/j.iswa.2024.200385.

A graph-based cardiac arrhythmia classification methodology using one-lead ECG recordings

Dorsa EPMoghaddam[a,*], **Ananya Muguli**[a], **Mehdi Razavi**[b], **Behnaam Aazhang**[a]

[a]Department of Electrical and Computer Engineering, Rice University, TX, United States of America

[b]Department of Cardiology, Texas Heart Institute, TX, United States of America

Abstract

In this study, we present a novel graph-based methodology for an accurate classification of cardiac arrhythmia diseases using a single-lead electrocardiogram (ECG). The proposed approach employs the visibility graph technique to generate graphs from time signals. Subsequently, informative features are extracted from each graph and then fed into classifiers to match the input ECG signal with the appropriate target arrhythmia class. The six target classes in this study are normal (N), left bundle branch block (LBBB), right bundle branch block (RBBB), premature ventricular contraction (PVC), atrial premature contraction (A), and fusion (F) beats. Three classification models were explored, including graph convolutional neural network (GCN), multi-layer perceptron (MLP), and random forest (RF). ECG recordings from the MIT-BIH arrhythmia database were utilized to train and evaluate these classifiers. The results indicate that the multi-layer perceptron model attains the highest performance, showcasing an average accuracy of 99.02%. Following closely, the random forest achieves a strong performance as well, with an accuracy of 98.94% while providing critical intuitions.

Keywords

Arrhythmia classification; Electrocardiogram (ECG); Graph convolutional neural network (GCN); Multi-layer perception (MLP); Random forest (RF); Visibility graph (VG)

1. Introduction

Cardiovascular diseases (CVDs), accounting for approximately 17.9 million deaths annually, are the leading cause of global mortality. Reducing the incidence of premature fatalities hinges on the accurate identification of individuals at the highest risk and ensuring their

*Corresponding author. de11@rice.edu (D. EPMoghaddam).

CRediT authorship contribution statement

Dorsa EPMoghaddam: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Ananya Muguli:** Conceptualization, Methodology, Software. **Mehdi Razavi:** Writing – review & editing, Funding acquisition. **Behnaam Aazhang:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

None Declared.

prompt access to appropriate treatment. Among the various manifestations of CVDs, arrhythmia, which denotes abnormal changes in the heart pulse rate, emerges as a common occurrence. A vital initial step in planning effective treatment for patients with cardiac problems is the establishment of a precise and automated arrhythmia diagnosis model. This pivotal tool plays a vital role in enhancing the accuracy of identification and expediting access to timely interventions, contributing to the overall goal of reducing the impact of cardiovascular diseases on global health.

The electrocardiogram (ECG) is a non-invasive medical diagnostic tool, providing essential insights into the heart's electrical activity, rhythm, and condition (Kligfield et al., 2007). It is widely employed by cardiologists as a primary assessment tool for evaluating patients. Traditionally, cardiologists invested significant time in manually reviewing electrocardiogram recordings. The integration of automated arrhythmia detection models can streamline this process, resulting in enhanced efficacy and accuracy in cardiac assessments. Usually, the 12-lead ECG obtained from 10 electrodes is used for a routine study of the electrical activity of the heart. These leads include three bipolar limb leads (I, II, III), three augmented unipolar limb leads (aVR, aVL, aVF), along with six precordial leads (V1–V6). From bipolar leads, the second lead is the most widely used as it generally provides more information on the important waves of a heartbeat Lai, Bu, Su, Zhang, and Ma (2020).

In recent years, the utilization of machine learning algorithms in analyzing diverse bio-signals has gained significant traction, demonstrating superior performance compared to previous methods (Ebrahimi, Loni, Daneshtalab, & Gharehbaghi, 2020; EPMoghaddam, Banta, Post, Razavi, & Aazhang, 2023; EPMoghaddam, Banta, et al., 2023a; EPMoghaddam, Sheth, Haneef, Gavvala, & Aazhang, 2022; Nasiri, Naghibzadeh, Yazdi, & Naghibzadeh, 2009; Wang, Chiang, Hsu, & Yang, 2013). Notably, numerous studies have concentrated on arrhythmia classification, with a discernible trend emphasizing the integration of neural network models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to augment the precision and efficacy of arrhythmia classification (Ebrahimi et al., 2020; EPMoghaddam, Muguli, & Aazhang, 2023b; Jun et al., 2018; Singh, Pandey, Pawar, & Janghel, 2018).

As an example, the study by Sahoo et al. introduces an algorithm designed for the detection of QRS complexes, which encompass three deflections on an ECG tracing: the Q wave, the R wave, and the S wave, alongside the extraction of features using multiresolution wavelet transform (Sahoo et al., 2017). These extracted features are subsequently employed for classifying cardiac abnormalities, including normal (N), left bundle branch block (LBBB), right bundle branch block (RBBB), and paced beats (P). The classification is carried out using neural network (NN) and support vector machine (SVM) classifiers. The reported average accuracy is 96.67% for NN and 98.39% for SVM. Another study investigates the application of continuous wavelet transform in comparison to alternative data transformation techniques, specifically discrete wavelet transform and discrete cosine transform, for the task of arrhythmia classification (Khorrami & Moavenian, 2010). This study employs multi-layer perceptron (MLP) and SVM as classifiers. Notably, the study utilizes a dataset comprising only ninety beats for each arrhythmia—a limited set employed for training, testing, and validation purposes. A more recent study in Houssein et al. (2021) explores the application

of manta ray foraging optimization with SVM for the efficient classification of ECG arrhythmias. This approach incorporates a diverse set of features, including one-dimensional local binary pattern, wavelet, higher-order statistical, and morphological information. The integration of these features into the proposed framework achieves an inter-patient average accuracy of 98.26%.

In the study by Jangra et al. a novel CNN model is proposed for arrhythmia classification, achieving an accuracy of 99.48% in distinguishing ventricular ectopic beats and supraventricular ectopic beats (Jangra, Dhull, Singh, Singh, & Cheng, 2023). Similarly, the work by Kim et al. (2023) introduces another CNN architecture named WavelNet for the same task. In their three-class classification, they attain sensitivity rates of 91.4%, 49.3%, and 91.4% for non-ectopic, supraventricular ectopic, and ventricular ectopic beat classifications, respectively. In the study presented in Madan et al. (2022), a hybrid model named 2D-CNN-LSTM is introduced, which combines convolutional neural network and long short-term memory (LSTM) network architectures. Initially, ECG signals are translated into scalogram images before being input into the model. The researchers utilize heartbeats from three databases and report accuracy rates of 98.7%, 99%, and 99% for the three-class classification of cardiac arrhythmias, congestive heart failure, and normal sinus rhythm, respectively.

Moreover, the visibility graph (VG) technique has been utilized in several studies for the arrhythmia classification task. For instance, in Li et al. (2021), weighted multi-scale limited penetrability visibility graph features were extracted, and an XGBoost classifier was employed to effectively classify atrial fibrillation (AF). Similarly, another study utilized the adjacency matrix of weighted visibility graphs or sequences of node weights as features for the classification of cardiac disorders using ECG recordings from the PTB-XL dataset (Kutluana & Türker, 2024). Table 1 summarizes several studies' strengths and limitations. As evident, there is still room for further investigation into a broader spectrum of arrhythmia classes and enhancing classification performance. Additionally, addressing the dataset's limitations, especially regarding data imbalance, is imperative as it can significantly affect learning outcomes. In this work, we tackle these gaps by examining a wider array of arrhythmia classes while concurrently enhancing classification accuracy. Moreover, we will discuss the dataset's constraints and propose viable strategies to mitigate any associated limitations.

In this study, we introduce and assess a novel methodology for performing the arrhythmia classification task by leveraging visibility graphs in conjunction with three classifiers: graph convolutional neural network (GCN), multi-layer perceptron, and random forest (RF). The paper is structured as follows: Section 2 delves into the data and provides a comprehensive methodology overview, while Section 3 presents the experimental results. Ultimately, Sections 4 and 5 discuss the findings and conclude the paper.

2. Material and methods

2.1. Data

The MIT-BIH open-source arrhythmia database from the PhysioNet Forum stands out as the most widely utilized dataset for the arrhythmia classification task (Goldberger et al., 2000; Xiao et al., 2023). In line with evaluating the presented model, we employed the MIT-BIH dataset, which encompasses 48 half-hour ECG recordings obtained from 47 distinct patients at the BIH laboratory. This dataset features 2-channel ECG recordings with a sampling rate of 360 Hz, offering a diverse range of heartbeats. For the specific focus of this work, we only utilize the first ECG lead, namely modified limb lead II (*MLII*). We narrowed down our analysis to data from six classes: N, LBBB, RBBB, PVC (premature ventricular), A (atrial premature contraction), and F (fusion beats). Fig. 1A illustrates the distribution of distinct types of heartbeats for each patient. Notably, recordings 102, 104, 207, and 217 were excluded from our analysis due to poor signal quality (Mark, 1987; Sellami & Hwang, 2019). Fig. 1B displays the proportion of various classes within the final training and testing sets.

2.2. Preprocessing

The first stage in the proposed methodology is preprocessing, focusing on noise removal and segmenting ECG recordings into individual heartbeats (Kher et al., 2019). Noise removal involves applying a notch filter to eliminate 60 Hz power line noise, followed by wavelet denoising using biorthogonal wavelets to effectively remove high-frequency noise (Mallat, 1999). Additionally, baseline wandering is addressed through the application of successive median filters.

Following the noise removal, ECG signals undergo a subsequent partitioning into individual heartbeats. In Fig. 2A, two ECG heartbeats and their essential components are displayed. A crucial consideration in this process was to ensure a uniform segment length, coupled with the centering of the R peaks across different segments. The goal was to have the same number of time samples in each heartbeat segment, ensuring that the peaks were consistently represented at roughly the same time across all segments. Opting for windows spanning 300 time points, equivalent to 0.83 s, was a deliberate choice to encompass the entire temporal scope of a beat, ensuring a comprehensive analysis that captures all critical components of a heartbeat. To accomplish this, as shown in Fig. 2B, for each heartbeat, the R peaks of the preceding and subsequent heartbeats are identified, and RR intervals are calculated. If, on each side, half of the RR interval is less than 150 time points, we apply padding to ensure a window of 300 centered around the R peak.

Following data segmentation, a down-sampling procedure was employed due to the substantial number of time points within each heartbeat. This down-sampling aimed to reduce the dimensionality of each segment, enhancing the efficiency of both graph generation and graph feature extraction processes. To achieve this, a sliding window approach was applied across the time-series, where the average of the amplitude values within each window was computed. Specifically, a sliding window of 10 time points was used, with a 20% overlap between consecutive windows. The sliding window duration

strikes a balance, being large enough to reduce complexity yet small enough to ensure that the final down-sampled segment, consisting of 38 time points, retains sufficient characteristics of the raw heartbeat without substantial distortion.

2.3. Visibility graph generation

The visibility graph is a technique used to transform complex time-series data into a network (Lacasa, Luque, Ballesteros, Luque, & Nuno, 2008; Stephen, Gu, & Yang, 2015). Fundamentally, each time point serves as a node in the graph, and the presence of an edge between two nodes signifies the intervisibility of the corresponding time points. In this study, we employ two variations of the visibility graph, specifically natural visibility graphs (NVG) and horizontal visibility graphs (HVG), as detailed in Lacasa et al. (2008), Luque, Lacasa, Ballesteros, and Luque (2009).

In the context of a natural visibility graph, two arbitrary points (t_a, y_a) and (t_b, y_b) possess visibility and are consequently connected if, for any other data point (t_c, y_c) situated between them, the following condition is satisfied:

$$y_c < y_b + (y_a - y_b) \times \frac{t_c - t_b}{t_a - t_b}. \tag{1}$$

In the case of HVG, two nodes are connected if:

$$y_c < y_b, y_a. \tag{2}$$

Fig. 3 provides an illustration of the process of creating an NVG from a heartbeat. For the conversion of time-series data to a visibility graph, the *ts2vg* library in Python is employed.

In the process of constructing a visibility graph, it is possible to incorporate edge weights. In our context, we introduced edge weights based on the distance between two points. Specifically, the weight assigned to the edge connecting two arbitrary points (t_a, y_a) and (t_b, y_b) is defined as:

$$w_{e_{ab}} = \sqrt{(y_a - y_b)^2 + (t_a - t_b)^2}, \tag{3}$$

where e_{ab} represents the edge connecting nodes a and b in the graph, and $w_{e_{ab}}$ is its corresponding weight.

The rationale for employing a visibility graph in the current task is rooted in the understanding that the peaks and their relative distances constitute the most informative aspects of a heartbeat. The success of this approach lies in the visibility graph's unique

capability to accentuate these critical features, aligning seamlessly with the essential requirements of the task.

In Fig. 4, we present an illustrative example featuring both types of graphs for a heartbeat. While HVGs generally exhibit fewer edges than NVGs due to a more stringent constraint on edge presence, leading to reduced computational costs, we find NVGs to be a more fitting choice. This preference arises because NVG places more focus and emphasis on the peaks, and given that heartbeat peaks play a crucial role in the diagnostic process, it aligns better with classification objectives.

To further appreciate the necessity of down-sampling as part of the preprocessing step, note that, initially, each heartbeat sample consisted of 300 time points. Creating a visibility graph directly from these samples would result in a graph with 300 nodes and numerous edges, imposing heavy computational demands for both graph creation and feature extraction. Instead, using down-sampled signals results in small graphs with 38 nodes.

2.4. Feature extraction

Three categories of features are extracted from each heartbeat sample. These features encompass graph features derived for each node in the visibility graph, temporal and entropy-based features extracted from 10-second sliding time windows, and time–frequency features.

The graph-based features include degree centrality, Katz centrality, betweenness centrality, closeness centrality, eigenvector centrality, Laplacian centrality, degree assortativity coefficient, and the graph's eccentricity. To extract these features, the Python library *NetworkX* was utilized (Hagberg, Swart, & S. Chult, 2008). The emphasis of VG on heartbeat peaks led us to choose centrality-based features. These features aim to characterize the connections among nodes in the graph and highlight prominent nodes representing the peaks of heartbeats.

In addition to these graph features, a set of time-domain features was computed for each small sliding window within a heartbeat during the down-sampling process. Subsequently, these features were associated with the corresponding node in the VG. These features encompass SVD entropy, approximate entropy, permutation entropy, the count of zero-crossings, Petrosian fractal dimension, Katz fractal dimension, Higuchi fractal dimension, and the energy of frequency sub-bands. Moreover, the short-time Fourier transform (STFT) was computed for each heartbeat, with window length and overlap matching the down-sampling step. Consequently, a vector of frequency content was assigned to each node in the graph. Ultimately, each node was represented by a feature vector comprising 22 elements. A number of previous studies have examined a subset of these features and highlighted their relevance.

Algorithm 1 Processing Flow

Input: Raw ECG recordings
Output: Visibility Graphs
for each $MLII$ **do**
  $y \leftarrow$ Apply notch filter to $MLII$
  $y \leftarrow$ Apply wavelet denoising to y
  $y \leftarrow$ Apply successive median filters to y
  for each Rin y **do**
    $W \leftarrow$ Select a window of 300 points around R
    $W \leftarrow$ Apply down-sampling to W
    $VG \leftarrow$ Create visibility graph from W
    $VG \leftarrow$ Extract nodes' features and add to VG
  end for
end for

2.5. Classification

Prior research has employed a variety of machine learning models for arrhythmia classification (Atal & Singh, 2020; Duong, Doan, Chu, & Nguyen, 2023; He et al., 2023; Zhang et al., 2021; Zhao, Liu, Han, & Peng, 2022). In this study, we employ three classifiers and assess their performance, including GCN, MLP, and RF. Fig. 5 provides an overview of the models employed in this context for the six-class arrhythmia classification task.

2.5.1. Graph convolutional neural network—The graph neural network (GNN) stands out as a potent neural network designed specifically for processing data structured as graphs. Several variations of GNNs, renowned for their robust performance in analyzing time-series data, have been extensively explored in prior studies (He et al., 2023; Zhao et al., 2022). One noteworthy category within GNNs is the graph convolutional neural network, which utilizes convolutional layers to gather information from neighboring nodes based on the connectivity within the graph. The number of layers in a GCN plays a crucial role in determining the extent to which neighboring nodes influence a given node.

2.5.2. Random forest—Random forest is a notable ensemble learning algorithm leveraging the bagging technique, which can be used in regression and classification tasks (Breiman, 2001). By constructing an ensemble of decision trees, this algorithm delivers a classification prediction determined by the mode of individual tree predictions. Renowned for their versatility and robustness, random forests excel in handling complex datasets. To employ an RF model, we concatenate all the graph nodes' features into a single feature vector representing that specific heartbeat sample. Consequently, each heartbeat is now represented by a vector of size $38 \times 22 = 836$, where 38 corresponds to the number of nodes in each graph, and 22 is the length of the feature vector for a single node. Fig. 5B depicts the outlined process.

2.5.3. Multi-layer perceptron—A multi-layer perceptron is a basic neural network architecture characterized by the inclusion of multiple fully connected layers. Our study explored the efficacy of MLPs with varying depths, specifically employing configurations with 2 to 5 layers.

The overall processing flow is outlined in Algorithm 1, detailing the steps for transforming raw ECG samples into visibility graphs. Following the completion of this stage, we proceed with the classification process. Algorithm 2 provides a comprehensive description of our classification approach, which utilizes the processed visibility graphs to predict arrhythmia classes.

Algorithm 2 Classification

Input: Visibility Graphs, Classifier
Output: Predicted Arrhythmia Classes
$VG_{\text{train}}, VG_{\text{test}} \leftarrow$ Divide visibility graphs into two sets
$Y_{\text{train}} \leftarrow$ Extract arrhythmia labels for VG_{train}
$Y_{\text{test}} \leftarrow$ Extract arrhythmia labels for VG_{test}
if Classifier is GCN **then**
    $X_{\text{train}} \leftarrow VG_{\text{train}}$
    $X_{\text{test}} \leftarrow VG_{\text{test}}$
else
    $X_{\text{train}} \leftarrow$ Extract feature vectors from VG_{train}
    $X_{\text{test}} \leftarrow$ Extract feature vectors from VG_{test}
end if
Train the Classifier on $(X_{\text{train}}, Y_{\text{train}})$
$Y_{\text{pred}} \leftarrow$ Predict labels for X_{test}
Evaluate the Classifier by comparing Y_{pred} and Y_{test}

3. Results

3.1. Performance measures

We investigate four widely used metrics, namely accuracy, precision, recall, and F1-score, to assess the performance of the proposed model. These metrics are defined as follows:

$$Accuracy = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} \tag{4}$$

$$Precision = \frac{t_p}{t_p + f_p} \tag{5}$$

$$Recall = \frac{t_p}{t_p + f_n} \tag{6}$$

$$F_1 = \frac{t_p}{t_p + \frac{1}{2}(f_p + f_n)} \tag{7}$$

Here t_p, f_p, t_n, f_n represent true positives, false positives, true negatives, and false negatives, respectively.

In many cases, solely examining one of these metrics can be misleading and may not furnish sufficient information for comprehensive assessments. For instance, in scenarios characterized by extreme class imbalance, a naive classifier might label every sample as the majority class, yielding a seemingly high accuracy. However, such a classifier is inherently undesirable. By incorporating the four aforementioned metrics, we ensure a thorough evaluation of the model's performance.

3.2. Results

We employ an intra-patient approach by consolidating samples from all patients. Following this amalgamation, we randomly divide the samples into two non-overlapping sets—one for training and the other for testing, with each set comprising 46,497 samples. The presented results are obtained by evaluating the performance of the trained models on the testing data.

In our exploration of classifiers, we assess three models: GCN, MLP, and RF. Each model undergoes an examination of a diverse range of parameters, including batch size, learning rate, the number of layers, the number of neurons in hidden layers, and more. Notably, for learning rate, values of 0.0001, 0.001, and 0.01 were explored, with 0.001 selected. These selections were made based on commonly used values for learning rates in various studies. Regarding batch size, we tested 32, 64, 128, and 256, finding that 32 and 64 achieved the highest performance.

For GCN models, we experimented with varying the number of layers from three to seven. Initially, we started with smaller models with fewer neurons and observed how incorporating additional layers and neurons impacted the performance. Table 2 presents GCN performance in several scenarios. Through our experimentation with three different layer configurations (3, 5, 7), we observed that employing seven layers resulted in a slightly better outcome, achieving an accuracy of 98.20%.

Table 3 showcases the performance of RF across various scenarios. Several choices for the number of trees in the forest, maximum tree depth, and minimum samples required to split an internal node are explored. In Table 3, unless specified otherwise, the parameter values are set to the defaults of the *RandomForestClassifier* function in *sklearn*. We also explored

the addition of class weights to address data imbalance but found it to be insufficient. Using RF classifiers, the highest accuracy achieved was 98.94%.

Table 4 showcases the performance of the proposed algorithm combined with the MLP classifier using various parameters. For MLPs, we initiated experiments with smaller network architectures comprising two to five layers. We employed configurations with 16, 32, 64, 128, 256, and 512 neurons and explored various combinations. Optimal performance was observed with an architecture comprising 3 layers and neuron counts of 128, 64, and 32, respectively. MLP exhibited slightly superior results compared to the other two classifiers, achieving an accuracy of 99.02%. Fig. 6 provides a detailed overview of the aforementioned classifier for each individual class. The first four classes exhibit higher accuracy, while the last two show slightly lower performance. This could be attributed, in part, to the lower number of samples associated with these classes in the training data, limiting the model's ability to learn their patterns effectively. Additionally, identifying fusion beats poses several challenges due to their complex morphology and variability. Fusion beats can manifest in diverse appearances, and the degree of fusion may vary among individuals. The diversity in morphology and timing makes it challenging for models to learn a universal representation of fusion beats, especially given the limited number of samples available. Furthermore, fusion beats may share characteristics with other beat types, such as PVC or normal beats, leading to overlap in features that complicates the model's ability to precisely differentiate between these classes (Das & Ari, 2014). Fig. 7 presents the accuracy and loss plots over 100 epochs for three models.

4. Discussion

Among the three classifiers, GCN demonstrates slightly lower performance. This observation may be attributed to the incorporation of graph-based features into the nodes' feature vectors and the subsequent aggregation of information from neighboring nodes in GCN, which may introduce redundancy, potentially diminishing the significance of aggregating information based on graph connectivity. Furthermore, the efficacy of aggregating neighbor information might be less pronounced because of the nature of features in comparison to scenarios with more elaborate relationships, such as social networks.

While the MLP stands out as the top performer, it is closely followed by RF. Notably, RF gains a distinct advantage by offering valuable insights into the significance of individual features. To gauge the importance of features, we can analyze the mean decrease in impurity. Features exhibiting a higher mean decrease in impurity hold greater importance in the classification process. Scrutinizing these prominent features offers valuable insights into the model's decision-making mechanisms. To enhance interpretability, it is insightful to correlate these features with their corresponding nodes in the graph and time points within the down-sampled signal. Given the construction of input features for the RF model, a mapping exercise is very straightforward and allows us to associate each feature with a specific node in the graph or a time point in the down-sampled signal, accompanied by its specific feature type. Such a contextualized analysis facilitates a deeper understanding of the model's reliance on specific nodes and characteristics, shedding light on the factors driving its classification decisions.

Analyzing the top 10 features underscores their predominant association with time points within the QRS area, notably excluding the R peak itself. In most instances of heartbeats, the R peak, serving as the highest point, demonstrates significant connectivity and thus displays limited variability across different target classes. Consequently, its utility in the classification process is diminished, as it fails to provide substantial discriminatory information. However, its neighboring nodes can provide insights into the width and slope of the QRS segment. Specifically, in a heartbeat sample with 38 nodes and node R located at time point 18, the top 10 features are derived from time points 16, 21, 22, and 23, encompassing node features such as Beta sub-band energy, degree centrality, and STFT features. Expanding our scrutiny to the top 20 features illuminates a consistent dominance of nodes 16, 17, 20, 21, 22, and 23, and feature categories such as Beta sub-band energy, degree centrality, STFT features, alpha sub-band energy, SVD entropy, and closeness centrality.

Diverging from the conventional reliance on intricate neural networks for feature extraction, our approach involves implementing an innovative methodology that constructs graph structures with a specific emphasis on the peaks of heartbeats—the most informative components in a heartbeat. Additionally, by incorporating STFT, we integrate frequency content into our analysis. This meticulous process results in a comprehensive and insightful feature vector that encapsulates crucial aspects of a heartbeat. All three models demonstrated comparably high performance, further highlighting the significance of the proposed features and their pivotal role in this classification task. It is noteworthy that when utilizing RF and MLP as classifiers, although we do not directly leverage the graph structure, we retain crucial insights into the visibility graph by utilizing the calculated graph features.

Table 5 offers a detailed comparison between our study and prior research on intra-patient ECG beat classification using the MIT-BIH database. Our proposed methodology outperforms many recent studies by leveraging a novel approach and achieving an accuracy of 99.02% on the classification of six arrhythmia classes (Kumar et al., 2023; Li et al., 2017; Oh et al., 2018; Wu et al., 2021). While certain studies may initially appear to showcase superior performance, it is imperative to acknowledge that they hinge upon considerably smaller subsets of recordings from the database (Shi et al., 2021; Yildirim, 2018; Yu & Chen, 2007). Using a limited sample size can potentially compromise the robustness and generalizability of research findings. Furthermore, it is noteworthy that most of these studies concentrate on five arrhythmia classes, whereas our case involves six classes, introducing additional complexity to the classification task.

A statistical method for comparing the performance of classifiers involves hypothesis testing. The p-value is a statistical measure that quantifies the strength of evidence against the null hypothesis and represents the probability of observing more extreme results if the null hypothesis were true. A small p-value indicates that the observed data is unlikely to occur if the null hypothesis were true, leading to the rejection of the null hypothesis in favor of the alternative hypothesis. Conversely, a large p-value suggests that the observed data is not unusual under the null hypothesis, providing insufficient evidence to reject the null hypothesis. In simpler terms, the p-value helps determine the likelihood of obtaining the observed results purely by chance, allowing researchers to assess the statistical significance of their findings. A common significance level for hypothesis testing is set at 0.05. In

this framework, the null hypothesis posits that two classifiers perform equally well. When comparing our classifier to the one studied in Pandey et al. (2023), the obtained p-value is 0.61. This value, calculated based on accuracy, precision, recall, and F_1 values, indicates that there is no compelling evidence to reject the null hypothesis. Hence, statistically speaking, there is no significant difference between the classifiers. However, it is worth noting that our study encompasses a larger sample size and a broader range of arrhythmia classes. Furthermore, compared to Wu et al. (2021) and Kumar et al. (2023), p-values of 0.08 and 0.18 were obtained, respectively, indicating similar outcomes. This comparison supports our study, as it performs similarly on a more challenging task (six-class classification) with a larger sample size.

One of the challenges we encountered in this study was the data imbalance, particularly concerning atrial premature contraction and fusion beats, as shown in Fig. 1B. Our exploration of incorporating sample weights to heighten the penalty for misclassifying these classes proved insufficient, signaling the need for additional measures to address this issue effectively. Consequently, we are considering the prospect of oversampling the minority class and undersampling the majority class as a viable solution to mitigate this challenge.

A limitation of this study is its intra-patient focus. Enhancing the reliability and generalizability of our classification model requires the accumulation of a more extensive array of samples for each arrhythmia class. Furthermore, the inclusion of a broader spectrum of patients and adopting an inter-patient paradigm—where the trained model undergoes testing on a cohort of entirely new patients not represented in the training process—are essential steps. This strategic expansion ensures a comprehensive evaluation of the model's performance across diverse patient populations, ultimately advancing its effectiveness and applicability. While the predominant approach in current studies leans towards the intra-patient paradigm, it is crucial to explore the inter-patient paradigm more extensively. The prevalence of the intra-patient paradigm in studies may be attributed to the challenges posed by the limited availability of samples for arrhythmic classes. The inadequacy of these samples poses a potential hindrance to the development of a comprehensive and generalized model. Exploring the inter-patient paradigm on a broader scale is integral to gaining a deeper understanding of arrhythmia classification dynamics and ensuring the model's adaptability across diverse patient cohorts.

Our research underscores the efficacy of visibility graphs in analyzing time-series data and providing rich insights, particularly in bio-signal analysis, and aligns with previous investigations into VG applications (Kutluana & Türker, 2024; Li et al., 2021). By further validating the utility of visibility graphs in arrhythmia classification, our findings contribute to establishing these representations as valuable tools in bio-signal analysis. Additionally, our research emphasizes the importance of exploring novel formats, such as graph representations, to extract meaningful insights from complex datasets, enhancing our understanding of bio-signals and paving the way for further advancements in this field.

Furthermore, our investigation identifies concerns regarding time complexity when dealing with large graphs and addresses them by integrating down-sampling techniques into our methodology. This approach aims to improve computational efficiency and enable smoother

analysis of large-scale datasets in future studies. Additionally, we recognize the limitation posed by data imbalance and recommend exploring advanced techniques to address this issue effectively, thereby ensuring more robust model performance. Besides, incorporating diverse datasets can mitigate the challenge of limited samples from minority classes and further enhance the model's reliability.

Our study contributes to enhancing the generalizability of arrhythmic heartbeat classification. Given the severe implications of arrhythmia, including cardiac arrests and mortality, detecting deviations from normal heart function is critical for preventing adverse outcomes. Leveraging the distinctive shape of a heartbeat through simple models based on graphs and graph features, was shown to be promising. Moving forward, we envision exploring the possibility of implementing this paradigm for online detection of arrhythmic heartbeats, potentially enabling real-time arrhythmia detection in patients and facilitating prompt intervention when necessary.

5. Conclusions

This study introduces an innovative technique for classifying arrhythmic ECG heartbeats, utilizing a graph-based methodology. The time-series data is transformed into graphs, from which informative features are extracted. Subsequently, the resulting graphs are fed into classifiers to determine the target heartbeat class. The chosen classifiers include graph convolutional networks, random forests, and multi-layer perceptrons. All three classifiers exhibit similar performance, with MLP demonstrating slightly superior results, showcasing a promising F1-score of 99%. Random forest achieved a close performance and also provided helpful insights on the features. The obtained result underscores the potential of the proposed methodology as an excellent model for computer-aided diagnosis of cardiac arrhythmias. For future directions, we aim to expand our research by incorporating a more extensive range of arrhythmia classes and involving a larger number of patients. Additionally, we aim to explore the inter-patient paradigm and investigate alternative classification models.

Acknowledgments

The research presented in this study is supported by grants from the National Institutes of Health, United States of America (R01HL144683-04) and the National Science Foundation, United States of America (CCF-1838873).

Data availability

Data will be made available on request.

References

Atal DK, & Singh M (2020). Arrhythmia classification with ECG signals based on the optimization-enabled deep convolutional neural network. Computer Methods and Programs in Biomedicine, 196, Article 105607. [PubMed: 32593973]

Breiman L (2001). Random forests. Machine Learning, 45, 5–32.

Das MK, & Ari S (2014). ECG beats classification using mixture of features. International Scholarly Research Notices, 2014.

Duong LT, Doan TT, Chu CQ, & Nguyen PT (2023). Fusion of edge detection and graph neural networks to classifying electrocardiogram signals. Expert Systems with Applications, 225, Article 120107.

Ebrahimi Z, Loni M, Daneshtalab M, & Gharehbaghi A (2020). A review on deep learning methods for ECG arrhythmia classification. Expert Systems with Applications: X, 7, Article 100033.

EPMoghaddam D, Banta A, Post A, Razavi M, & Aazhang B (2023). Reconstructing 12-lead surface electrocardiogram from reduced lead sets: An encoder-decoder convolutional neural network approach. Available at SSRN 4535793.

EPMoghaddam D, Banta A, Post A, Razavi M, & Aazhang B (2023a). A novel method for 12-lead ECG reconstruction. In 2023 57th asilomar conference on signals, systems, and computers (pp. 1054–1058). IEEE.

EPMoghaddam D, Muguli A, & Aazhang B (2023b). A novel cardiac arrhythmia classification method using visibility graphs and graph convolutional network. In 2023 57th asilomar conference on signals, systems, and computers (pp. 866–870). IEEE.

EPMoghaddam D, Sheth SA, Haneef Z, Gavvala J, & Aazhang B (2022). Epileptic seizure prediction using spectral width of the covariance matrix. Journal of Neural Engineering, 19(2), Article 026029.

Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. circulation, 101(23), e215–e220. [PubMed: 10851218]

Hagberg A, Swart P, & Chult DS, (2008). Exploring network structure, dynamics, and function using NetworkX: Tech. rep, Los Alamos, NM (United States): Los Alamos National Lab.(LANL).

He Z, Chen Y, Yuan S, Zhao J, Yuan Z, Polat K, et al. (2023). A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ECG classification. Expert Systems with Applications, 221, Article 119711.

Houssein EH, Ibrahim IE, Neggaz N, Hassaballah M, & Wazery YM (2021). An efficient ECG arrhythmia classification method based on Manta ray foraging optimization. Expert Systems with Applications, 181, Article 115131.

Jangra M, Dhull SK, Singh KK, Singh A, & Cheng X (2023). O-WCNN: An optimized integration of spatial and spectral feature map for arrhythmia classification. Complex & Intelligent Systems, 9(3), 2685–2698. [PubMed: 34777963]

Jun TJ, Nguyen HM, Kang D, Kim D, Kim D, & Kim Y-H (2018). ECG arrhythmia classification using a 2-D convolutional neural network. arXiv:1804.06812 [cs.CV].

Kher R, et al. (2019). Signal processing techniques for removing noise from ECG signals. Journal of Biomedical Engineering and Research, 3(101), 1–9.

Khorrami H, & Moavenian M (2010). A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. Expert Systems with Applications, 37(8), 5751–5757.

Kim N, Seo W, Kim J.-h., Choi SY, & Park S-M (2023). WavelNet: A novel convolutional neural network architecture for arrhythmia classification from electrocardiograms. Computer Methods and Programs in Biomedicine, 231, Article 107375. [PubMed: 36724593]

Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, et al. (2007). Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the American college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology. Circulation, 115(10), 1306–1324. [PubMed: 17322457]

Kumar S, Mallik A, Kumar A, Del Ser J, & Yang G (2023). Fuzz-ClustNet: Coupled fuzzy clustering and deep neural networks for arrhythmia detection from ECG signals. Computers in Biology and Medicine, 153, Article 106511. [PubMed: 36608461]

Kutluana G, & Türker İ (2024). Classification of cardiac disorders using weighted visibility graph features from ECG signals. Biomedical Signal Processing and Control, 87, Article 105420.

Lacasa L, Luque B, Ballesteros F, Luque J, & Nuno JC (2008). From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences, 105(13), 4972–4975.

Lai D, Bu Y, Su Y, Zhang X, & Ma C-S (2020). Non-standardized patch-based ECG lead together with deep learning based algorithm for automatic screening of atrial fibrillation. IEEE Journal of Biomedical and Health Informatics, 24(6), 1569–1578. [PubMed: 32175879]

Li W, Wang H, Zhuang L, Han S, Zhang H, & Wang J (2021). Weighted multi-scale limited penetrable visibility graph for exploring atrial fibrillation rhythm. Signal Processing, 189, Article 108288.

Li D, Zhang J, Zhang Q, & Wei X (2017). Classification of ECG signals based on 1D convolution neural network. In 2017 IEEE 19th international conference on e-health networking, applications and services (pp. 1–6). IEEE.

Luque B, Lacasa L, Ballesteros F, & Luque J (2009). Horizontal visibility graphs: Exact results for random time series. Physical Review E, 80(4), Article 046103.

Madan P, Singh V, Singh DP, Diwakar M, Pant B, & Kishor A (2022). A hybrid deep learning approach for ECG-based arrhythmia classification. Bioengineering, 9(4), 152. [PubMed: 35447712]

Mallat S (1999). A wavelet tour of signal processing. Elsevier.

Mark R (1987). AAMI-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms. Association for the Advancement of Medical Instrumentation, Arrhythmia Monitoring Subcommittee, AAMI ECAR, 1987.

Midani W, Ouarda W, & Ayed MB (2023). DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals. Biomedical Signal Processing and Control, 85, Article 104954.

Nasiri JA, Naghibzadeh M, Yazdi HS, & Naghibzadeh B (2009). ECG arrhythmia classification with support vector machines and genetic algorithm. In 2009 third UKsim European symposium on computer modeling and simulation (pp. 187–192). IEEE.

Oh SL, Ng EY, San Tan R, & Acharya UR (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in Biology and Medicine, 102, 278–287. [PubMed: 29903630]

Pandey SK, Shukla A, Bhatia S, Gadekallu TR, Kumar A, Mashat A, et al. (2023). Detection of arrhythmia heartbeats from ECG signal using wavelet transform-based CNN model. International Journal of Computational Intelligence Systems, 16(1), 80.

Sahoo S, Kanungo B, Behera S, & Sabut S (2017). Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities. Measurement, 108, 55–66.

Sellami A, & Hwang H (2019). A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. Expert Systems with Applications, 122, 75–84.

Shi Z, Yin Z, Ren X, Liu H, Chen J, Hei X, et al. (2021). Arrhythmia classification using deep residual neural networks. Journal of Mechanics in Medicine and Biology, 21(10), Article 2140067.

Singh S, Pandey SK, Pawar U, & Janghel RR (2018). Classification of ECG arrhythmia using recurrent neural networks. Procedia Computer Science, 132, 1290–1297.

Stephen M, Gu C, & Yang H (2015). Visibility graph based time series analysis. PLoS One, 10(11), Article e0143015. [PubMed: 26571115]

Wang J-S, Chiang W-C, Hsu Y-L, & Yang Y-TC (2013). ECG arrhythmia classification using a probabilistic neural network with a feature reduction method. Neurocomputing, 116, 38–45.

Wu M, Lu Y, Yang W, & Wong SY (2021). A study on arrhythmia via ECG signal classification using the convolutional neural network. Frontiers in Computational Neuroscience, 14, Article 564015. [PubMed: 33469423]

Xiao Q, Lee K, Mokhtar SA, Ismail I, Pauzi A. L. b. M., Zhang Q, et al. (2023). Deep learning-based ECG arrhythmia classification: A systematic review. Applied Sciences, 13(8), 4964.

Yildirim Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine, 96, 189–202. [PubMed: 29614430]

Yu S-N, & Chen Y-H (2007). Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Pattern Recognition Letters, 28(10), 1142–1150.

Zhang H, Liu C, Zhang Z, Xing Y, Liu X, Dong R, et al. (2021). Recurrence plot-based approach for cardiac arrhythmia classification using inception-ResNet-v2. Frontiers in Physiology, 12, Article 648950. [PubMed: 34079470]

Zhao X, Liu Z, Han L, & Peng S (2022). Ecgnn: Enhancing abnormal recognition in 12-lead ecg with graph neural network. In 2022 IEEE international conference on bioinformatics and biomedicine (pp. 1411–1416). IEEE.

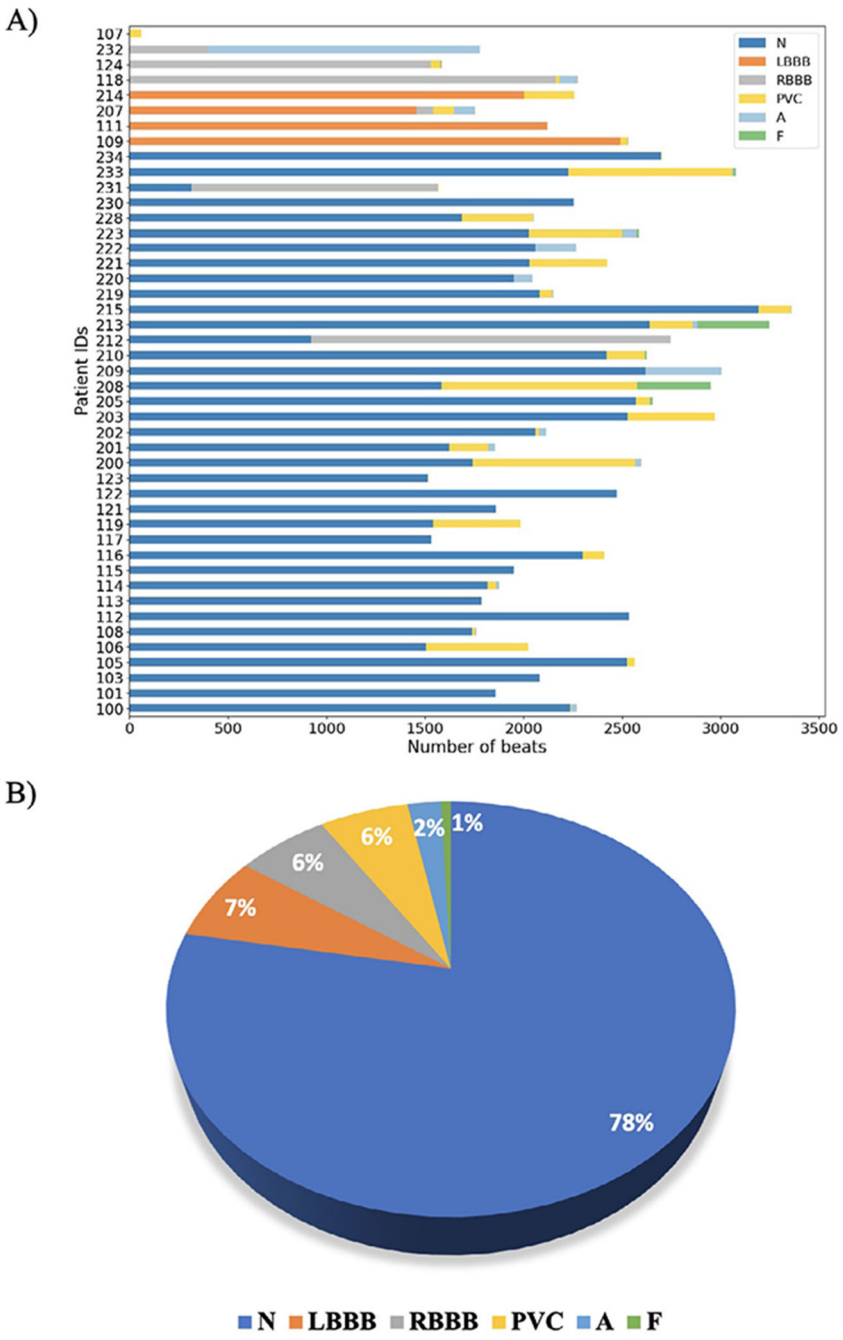

Fig. 1.

(A) Distribution of different heartbeat types for individuals in MIT-BIH database. (B) Data distribution in both training and testing sets. The dataset exhibits a significant imbalance, with the majority of instances belonging to the normal heartbeat class. This leads to challenges in training a model that adequately captures patterns in the minority classes.

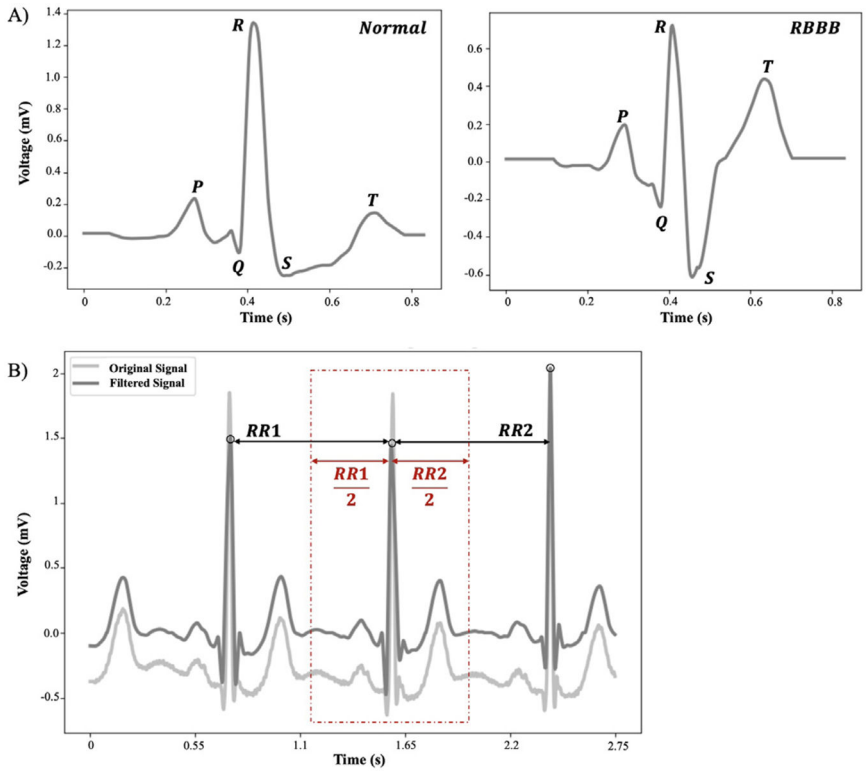

Fig. 2.

(A) Two heartbeat samples and their components. The plot on the right illustrates a normal heartbeat, while the one on the left represents a right bundle branch block heartbeat. (B) An example of an ECG signal before and after noise removal is shown. Note that black circles indicate R peaks. After noise removal, the signal is partitioned into individual heartbeat samples. Each sample has a fixed length of 300 time points. To achieve this, the R peaks of the preceding and subsequent heartbeats are identified, and RR intervals are calculated. On each side of the main heartbeat, if half of the RR interval is less than 150 time points, we apply padding to ensure a window of 300 centered around the R peak.

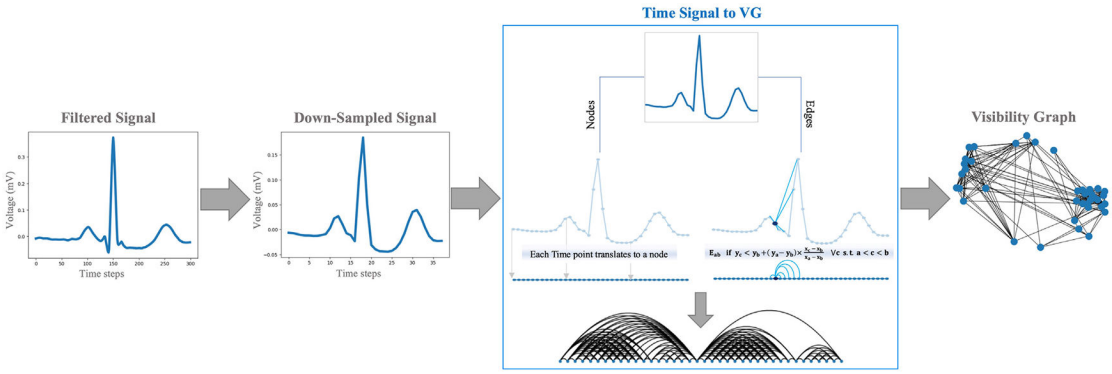

Fig. 3.
Natural visibility graph technique.

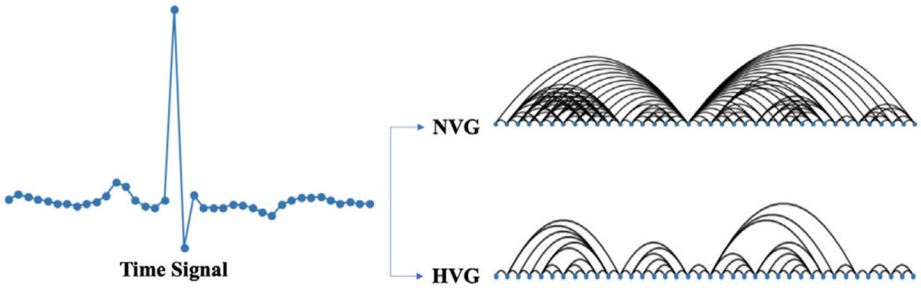

Fig. 4.
An example of NVG and HVG created from the same heartbeat.

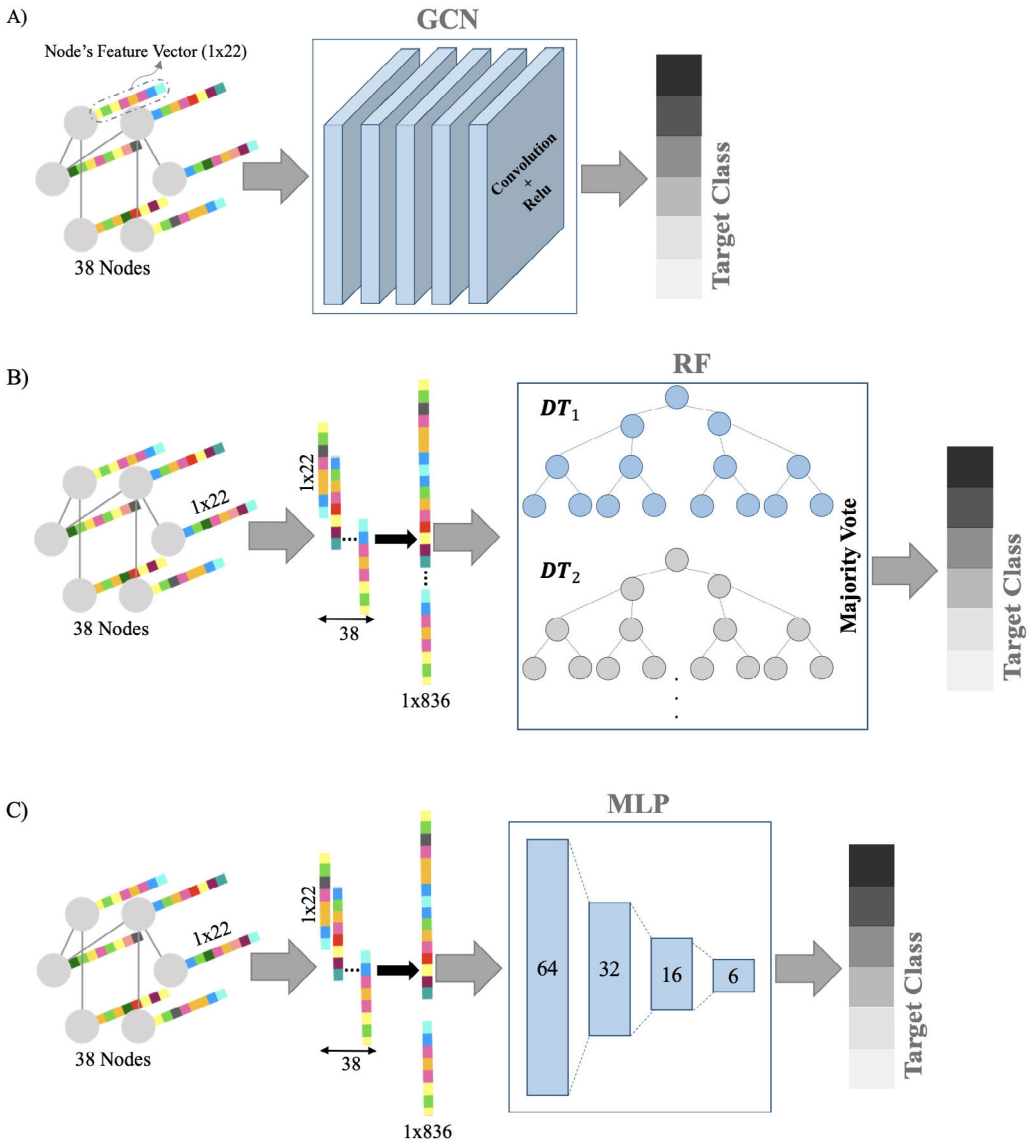

Fig. 5.

Three classifiers used for arrhythmia classification task. (A) Graph convolutional network. The GCN processes natural visibility graphs as input, each comprising 38 nodes. Each node is represented by a feature vector of size 1 × 22. The GCN then produces output corresponding to one of the six arrhythmia classes. (B) Random forest. Please note that DT stands for decision tree. The features of graph nodes are concatenated into a single vector, which is then inputted into the classifier. (C) Multi-layer perceptron.

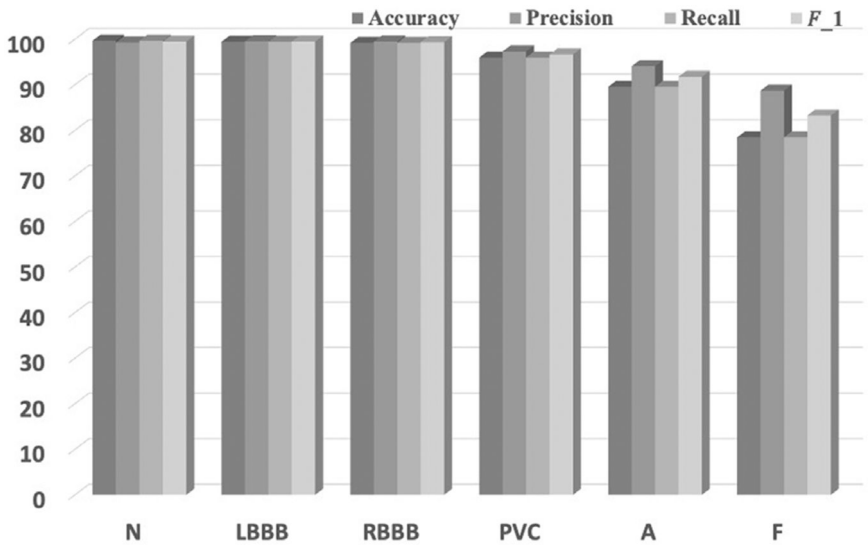

Fig. 6.
Classification details for the model achieving the highest performance.

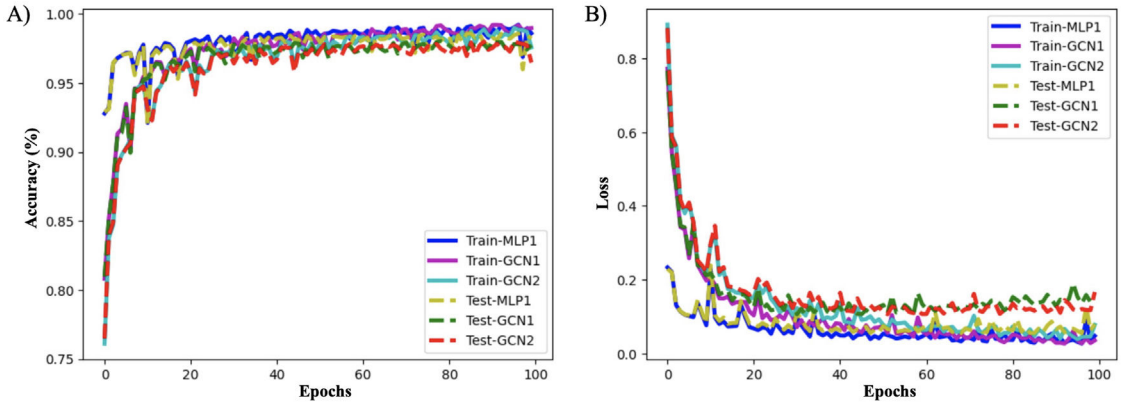

Fig. 7.

Accuracy and loss curves for three classifiers. MLP1 is an MLP with three layers (64, 32, 16), GCN1 is a 5-layer GCN with 150 neurons, and GCN2 is a 5-layer GCN with 100 neurons. (A) Accuracy plot. (B) Loss plot.

Table 1

Detailed comparison of several previous studies and their strengths and limitations.

Study	#Classes	Method	Strength	Limitations
Sahoo, Kanungo, Behera, and Sabut (2017)	4	Introduces an algorithm for QRS complex detection and feature extraction using multiresolution wavelet transform.	Intuitive features, Utilizing multiple classifiers, High accuracy in classification.	Limited number of arrhythmia classes, Lower number of samples, Limited discussion on dataset size and imbalance.
Khorrami and Moavenian (2010)	5	Compares continuous wavelet transform with two other techniques for arrhythmia classification.	Exploring the impact of employing two ECG input leads on classification performance in contrast to using a single lead.	Limited number of arrhythmia classes, Limited data size, Lack of discussion on dataset size and imbalance.
Houssein, Ibrahim, Neggaz, Hassaballah, and Wazery (2021)	4	Investigates the application of manta ray foraging optimization with SVM for arrhythmia classification.	High inter-patient accuracy, Comparison with several other metaheuristic algorithms.	Lack of discussion on the effect of data imbalance on the training.
Kim, Seo, Kim, Choi, and Park (2023)	3	Introduces a CNN architecture named WavelNet for arrhythmia classification.	Inter-patient paradigm, Addressing imbalance by incorporating class weights.	Model's dependence on the choice of mother wavelets, Moderate performance.
Li et al. (2021)	4	Offers a visibility-graph based denoising approach followed by XGboost to identify AF patients.	Novel denoising approach, Diverse dataset.	Focused only on AF, Lack of discussion on dataset imbalance.
Kutluana and Türker (2024)	5	Uses adjacency matrix of weighted visibility graphs for arrhythmia classification.	Exploring graph representation, Multi-label classification, Inter-patient paradigm.	Using 12-lead ECG as input, Moderate performance.

Table 2

GCN performance. Please note that for precision, recall, and F_1 columns, the first value represents the macro average, while the second one represents the weighted average. Learning rate is 0.001 and batch size is 64.

#Hidden layers	#Neurons	Class weight	Accuracy%	Precision%	Recall%	F_1 %
3	100	–	97.59	91.86, 97.55	87.72, 97.59	89.62, 97.52
3	150	–	97.89	93.45, 97.83	87.61, 97.89	90.19, 97.83
5	100	–	98.12	94.71, 98.07	89.07, 98.12	91.58, 98.07
5	150	–	98.13	93.16, 98.12	90.45, 98.12	91.59, 98.10
5	150	[0.5, 1, 1, 1, 1, 1]	98.03	98.95, 97.97	89.35, 98.03	91.43, 97.98
7	100	–	**98.20**	94.05, 98.17	90.60, 98.20	92.14, 98.18

Table 3

RF performance with different parameters. Please note that for precision, recall, and F_1 columns, the first value represents the macro average, while the second one represents the weighted average.

#Estimators	Max depth	Min samples split	Class weight	Accuracy%	Precision%	Recall%	F_1 %
20	–	2	–	98.39	97.75, 98.38	87.28, 98.39	91.61, 98.31
50	–	2	[0.1, 1, 1, 1.5, 3, 10]	98.32	98.38, 98.33	87.23, 98.32	91.89, 98.24
100	–	5	–	98.52	98.15, 98.51	87.80, 98.52	92.09, 98.44
100	10	5	[1, 1, 1, 1, 5, 20]	97.95	95.74, 97.94	89.25, 97.95	92.31, 97.90
150	–	5	–	98.51	98.16, 98.51	87.95, 98.51	92.21, 98.43
150	10	2	–	97.82	97.92, 97.84	82.89, 97.82	88.45, 97.67
200	–	2	–	98.55	97.99, 98.55	88.45, 98.56	92.50, 98.49
200	–	10	[1, 1, 1, 1, 5, 20]	98.52	97.82, 98.51	88.93, 98.52	92.81, 98.46
300	–	2	–	98.56	97.97, 98.55	88.17, 98.56	92.28, 98.48
300	–	10	–	98.48	98.24, 98.48	87.73, 98.48	92.08, 98.41
400	–	–	–	**98.94**	98.15, 98.93	91.70, 98.94	94.60, 98.91

Table 4

MLP performance. Please note that for precision, recall, and F_1 columns, the first value represents the macro average, while the second one represents the weighted average. Learning rate is 0.001 and batch size is 32.

#Hidden layers	#Neurons	Accuracy%	Precision%	Recall%	F_1 %
2	64, 32	98.67	95.55, 98.64	91.01, 98.67	93.05, 98.64
3	64, 32, 16	98.62	95.82, 98.59	90.79, 98.64	93.09, 98.59
3	128, 64, 32	**99.02**	95.92, 99.00	94.07, 99.02	94.97, 99.01
4	128, 64, 32, 16	98.78	95.96, 98.75	92.52, 98.78	94.14, 98.76
5	512, 256, 128, 64, 64	98.74	96.06, 98.71	92.02, 98.74	93.91, 98.71

Table 5

Performance comparison between the proposed work and previous studies using MIT-BIT database. All the studies are performing the intra-patient paradigm.

Study	ECG beat classes	#Records	#ECG samples	Method	Accuracy%
Yu and Chen (2007)	N, L, R, V, A, P	23	23,200	Probabilistic Neural Network	99.65
Li, Zhang, Zhang, and Wei (2017)	N, L, R, V, A	44	26,400	1D-CNN	97.50
Oh, Ng, San Tan, and Acharya (2018)	N, L, R, V, A	–	16,499	CNN+LSTM	98.10
Yildirim (2018)	N, L, R, V, P	–	7,326	ULSTM and BLSTM	99.39
Shi et al. (2021)	N, L, R, V, A	23	9,943	1D-CNN	99.59
Wu, Lu, Yang, and Wong (2021)	N, L, R, V, A	16	32,422	1D-CNN	97.20
Midani, Ouarda, and Ayed (2023)	N, L, R, V, A	48	100,062	DeepArr	99.46
Kumar, Mallik, Kumar, Del Ser, and Yang (2023)	N, S, V, F, Q[a]	44	87,554	Fuzz-ClustNet	98.66
Pandey et al. (2023)	N, L, R, V, A	–	25,000	1D-CNN	99.40
Proposed Work	N, L, R, V, A, F	44	92,994	VG+MLP	**99.02**

[a]AAMI recommended arrhythmia classes.