ECG Abnormality Detection via Visibility Graph-Based Learning: A Step Toward Effective and Explainable AI in Cardiology

No Author Given

Abstract. The automatic classification of electrocardiograms (ECG) signals is a well-established field in machine learning, yet conventional deep learning approaches primarily rely on sequential or convolutional methods that often fail to capture the intricate morphologic and structural dependencies within cardiac signals. In this work, we introduce a novel graph-based ECG classification framework that transforms raw time series data into structured visibility graphs, allowing for a richer and more interpretable representation of the heart's electrical activity. By utilizing a Graph Neural Network (GNN) architecture that integrates Graph Attention Networks (GATs) and Graph Isomorphism Networks (GINs), our approach brings new perspectives to traditional modeling, with promising relation-based design. The framework also provides a pathway to investigate model interpretability, enhancing clinical understanding of cardiac anomalies while maintaining diagnostic precision. These capabilities lay the groundwork for advancing explainable AI applications in cardiology. Our results demonstrate superior recall and F1-score compared to baseline methods, underscoring the potential of morphology graph-based models in medical signal analysis, particularly for cardiac abnormality detection.

Keywords: Electrocardiogram (ECG) · Graph Neural Network (GNN) · Visibility Graph · Classification

1 Introduction

Electrocardiography (ECG) is a fundamental non-invasive diagnostic tool for assessing cardiac electrical activity. By placing 12 electrodes on the patient's body, an ECG captures cardiac signals from multiple perspectives, providing a complete depiction of cardiac function. Analyzing these signals is crucial for detecting arrhythmias, ischemia, and other cardiovascular anomalies, enabling early intervention and improving patient outcomes. Given ECG's clinical importance and inherent complexity, automatic classification has received significant attention in computational research, focusing on enhancing diagnostic accuracy and clinical applicability.

State-of-the-art ECG classification methods predominantly employ deep learning architectures—such as Convolutional Neural Networks (CNNs) [6] and attention-based models [18]—to process ECG data as sequential numerical vectors. Despite their success, CNNs can struggle with long-range temporal dependencies, while attention mechanisms, though effective for global context, can

be computationally expensive and prone to noise sensitivity. Both often impose fixed feature extraction pipelines that may not fully capture intricate temporal and relational characteristics.

Beyond algorithmic limitations, ECG classification also faces challenges in data representation. Although thorough, 12 leads can include overlapping projections of the same cardiac events, increasing computational demands. Recent research shows that selective lead subsets can retain diagnostic accuracy while reducing complexity, particularly for wearable or resource-constrained settings [5]. Moreover, high-dimensional ECG data further elevates model complexity, underscoring the importance of eliminating redundant information while preserving key diagnostic elements.

To address these issues, we adopt a visibility graph representation combined with lead optimization. Visibility graph transformations convert ECG signals into network structures whose nodes and edges preserve local morphology and global rhythm patterns. In parallel, we leverage lead D2 robust rhythm information, augmented by morphological features from leads D1, V1, and V6, striking a balance between computational efficiency and clinical interpretability. As illustrated in Figure 1, our pipeline processes pre-filtered ECG signals into visibility graphs, allowing more efficient and intuitive abnormality classification.

,

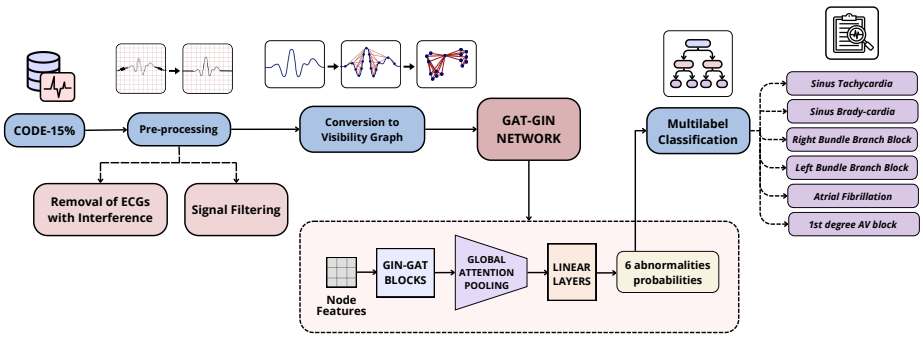

Fig. 1. ECG classification pipeline. The CODE-15% dataset undergoes preprocessing, visibility graph conversion, and feature extraction. A GAT-GIN network processes the graphs, generating probability scores for six cardiac abnormalities.

Building on this graph-theoretic foundation, we developed a GNN-based classification framework that directly analyzes visibility graph structures. Specifically, we integrate Graph Attention Networks (GATs) [14] and Graph Isomorphism Networks (GINs) [16] in an architecture designed to capture both structural and relational features. The GAT layers apply adaptive attention to emphasize clinically relevant interactions, while the GIN layers distinguish structurally similar but clinically distinct waveforms through localized node representations.

This structured representation provides an interpretable diagnostic pathway, enabling clinicians to cross-reference AI outputs with standard electrocardiographic markers. For instance, altered edge patterns at ST-segment nodes can highlight pathological changes, mitigating black-box concerns and fostering trust in AI diagnostics. Preliminary evaluations indicate performance on par with conventional methods, with the added benefit of graph-based interpretability. Physicians can visually trace waveform alterations—like ST-segment shifts—through the network topology, bridging the gap between computational analysis and clinical reasoning.

By leveraging this structured and attentional design, the model refines feature extraction, capturing both fine-grained local variations and broader topological dependencies. The result is a powerful framework capable of robust classification across diverse ECG patterns, opening new possibilities for model explainability, improving reliability and fostering clinical acceptance.

We validate our approach in a set of over 300k ECG exams to identify six clinically significant abnormalities: right bundle branch block (RBBB), left bundle branch block, atrial fibrillation, sinus tachycardia, sinus bradycardia, and first-degree atrioventricular block (1dAVB), characterized by distinct rhythm and morphological waves. We obtained results of f1-score of 0.815, being competitive with other state-of-the-art methods while excelling in detecting RBBB and 1dAVB, the latter being extremely difficult to be detected by automatic methods.

2 Related Works

2.1 ECG Classification

Automatic ECG classification has improved significantly with the advancement of deep learning models, particularly CNNs. Their ability to automatically extract discriminative features from raw ECG signals has greatly reduced the reliance on manual feature engineering. Several studies have demonstrated that CNN-based models can achieve cardiologist-level performance in arrhythmia detection, enhancing their potential for clinical application [9]. These models have been particularly successful in detecting heart arrhythmia, which can otherwise be challenging for traditional machine learning techniques.

However, CNNs face limitations in modeling long-range temporal dependencies in ECG signals, which are essential for identifying abnormalities such as intermittent arrhythmias. Capturing these long-term patterns in cardiac signals is crucial for improving diagnostic accuracy [13]. Furthermore, conventional methods tend to treat ECG signals as one-dimensional sequences, potentially overlooking more complex structural relationships within the waveform. Ultimately, noise contamination and signal variability are known to degrade the performance of deep learning models, making the robustness of such models in noisy environments a critical area for amplification.

To overcome such challenges, graph-based approaches have emerged as a promising alternative. By transforming ECG signals into graph representations,

these methods can preserve both local relationships between waveform peaks and valleys and the global structural dependencies within the signal. This approach offers greater flexibility in feature extraction, potentially improving the robustness and accuracy of automated ECG classification. Moreover, GNNs stand out for their possibilities of capturing both local and global features hierarchically, providing a more interpretable approach to ECG signal analysis [17].

2.2 Graph-Based Representations for Time Series Data

As an alternative to traditional sequence-based models, graph-based representations of time series leverage the relationships between vertices to encode both temporal and structural patterns. Visibility Graphs (VG) [8] are among the earliest approaches to transforming time series into graph structures, based on geometric visibility criteria, making them particularly suited for capturing complex dependencies in ECG signals.

Leveraging these time series representations, a range of algorithms and models have been uncovered, such as Temporal Graph Convolution Networks (T-GCN) [19] and Multivariate Time Graph Neural Networks (MTGNN) [15]. Moreover, shapelet-based algorithms like Time2Graph and Time2Graph+ [2] concatenate attention mechanisms to identify discriminative temporal features in evolving graph structures.

A more recent approach, Graph Neural Alchemist (GNA), introduces a modular architecture that represents time series as directed visibility graphs, incorporating structural properties such as in-degree and PageRank metrics. This method demonstrates robustness and strong performance in classification tasks across multiple datasets [3]. Although indicating feasibility, these graph-based approaches remain underexplored for ECG classification.

3 Method

3.1 From ECG Time Series to Visibility Graphs

A core component of our work is the transformation of ECG signals into a graph-based representation using the visibility graph (VG) framework. In this approach, each temporal point in the ECG is mapped to a node, and edges are formed based on visibility criteria—two points are connected if a straight line between them does not intersect any intermediate points.

The graph construction described and illustrated in Figure 2 encodes the topographic and morphological relationships between ECG points. For instance, subtle morphological differences between ECG signals, such as variations in waveform shapes, can obstruct the visibility between nodes in one condition while preserving visibility in another, effectively distinguishing them in the resulting graphs. Consequently, abnormalities typically yield distinct connectivity patterns, enhancing the discriminative capability of the graph representation.

To balance efficiency with clinical relevance, our approach centers on constructing visibility graphs from lead II (D2), the primary lead used for rhythm

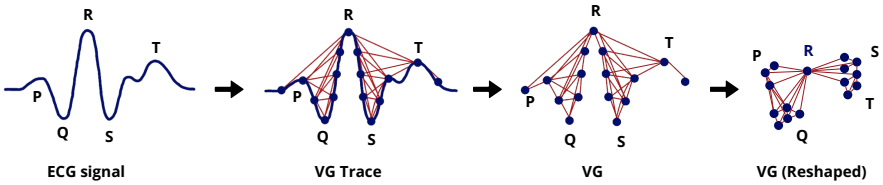

Fig. 2. Diagram showcasing the transformation of an ECG signal into a VG

assessment by cardiologists. This ensures that the graph structure accurately encodes the morphology of key waveforms, including the P wave, QRS complex, and T wave. However, to enhance the discriminative capability while avoiding excessive computational overhead, we augment node features with data from three additional leads: D1, V1, and V6. The selection of these specific derivations was guided by clinical reasoning and validated by a cardiologist, given their ability to capture distinct and complementary aspects of cardiac activity.

Therefore, each node in the visibility graph corresponds to a temporal point in the ECG signal from lead D2, and is enriched with features derived from the selected leads D1, D2, V1, and V6. These features include degree, amplitude, and derivative, chosen to capture key aspects of ECG morphology. The feature extraction is described in Figure 3.

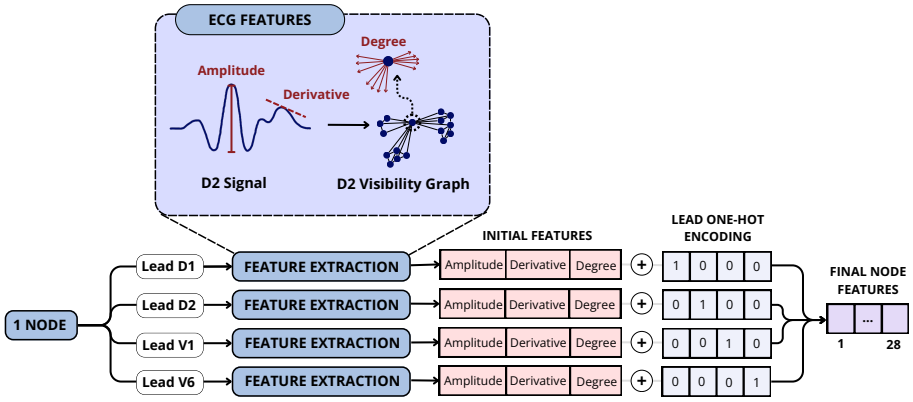

Fig. 3. Feature extraction pipeline for ECG signals. The D2 signal is converted into a Visibility Graph (VG), from which amplitude, derivative, and degree features are extracted for each lead (D1, D2, V1, and D1). These features are concatenated with a one-hot encoding of the lead, forming a 28-dimensional node feature vector.

Firstly, the degree quantifies the connectivity of each point in the ECG visibility graph, emphasizing the structural relevance of specific waveform components. This is particularly important because key points, such as the peaks of the P

wave, the QRS complex, and the T wave, exhibit distinct connectivity patterns that are critical for detecting pathological conditions. A notable example is Right Bundle Branch Block (RBBB), which modifies the QRS complex by introducing a secondary R wave in leads V1-V2. This alteration significantly reshapes the visibility graph, increasing the degree of secondary peaks and modifying connectivity patterns in a way that is highly distinctive for RBBB identification.

Secondly, the amplitude represents the raw electrical potential of the signal, serving as a fundamental parameter for differentiating ECG wave components. The QRS complex amplitude directly corresponds to ventricular depolarization strength, while P wave amplitude provides insights into atrial activity. In conditions such as ST-segment deviations, amplitude variations serve as critical indicators of myocardial infarction or ischemia. Notably, significant ST-segment elevation or depression signals ischemic events, making this feature essential for capturing diagnostically relevant information.

Finally, the signal derivative is employed to detect rapid variations during depolarization and repolarization phases, which are crucial for identifying arrhythmias. In Atrial Fibrillation (AF), the absence of well-defined P waves and highly irregular RR intervals introduce erratic fluctuations in the ECG waveform. These abrupt changes are captured by the derivative, which highlights sudden variations and serves as a powerful tool for distinguishing AF from other arrhythmias.

To ensure that the model differentiates between sources of information, each feature is accompanied by one-hot encoded vectors indicating the lead of origin (D1, D2, V1, V6). This results in a 28-dimensional feature vector per node, effectively encoding both the cardiac signal characteristics and their respective sources.

This feature structure allows the model to capture both local variations in ECG morphology and global dependencies between points, resulting in a rich and accurate representation of cardiac activity. The use of four selected leads and the inclusion of these features provides an efficient and effective representation of the ECG, balancing clinical detail with computational efficiency and lead optimizations. Additionally, the representativeness of the ECG signal is preserved, facilitating clinical interpretability, as the model can distinguish the structural and morphological differences between normal and pathological patterns.

To mitigate the effects of disparities in the scales of different features, each column of the 28-dimensional vector is standardized. Specifically, for the amplitude, derivative, and degree features, the mean (μ) and standard deviation (σ) are calculated using the valid nodes across all processed exams using the following transformation:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma},$$

where x represents the original value of a feature. This standardization ensures that each feature has zero mean and unit variance, which is crucial for stabilizing the model training and ensuring a balanced contribution from all variables during the learning process.

The final representation not only enhances the model's efficiency and expressiveness but significantly improves its clinical interpretability. In a medical context, visibility graphs intuitively capture morphological and topographical variations in cardiac signals. Normal ECG patterns produce well-structured, highly connected graphs, whereas abnormalities lead to more fragmented connectivity patterns. This distinction makes it easier to differentiate between healthy and pathological rhythms, providing greater transparency for healthcare professionals. Moreover, our approach strikes a balance between computational feasibility, diagnostic accuracy, and clinical interpretability—key aspects for practical applications in cardiology.

3.2 Graph Neural Network Architecture

The proposed network employs innovative GNNs to exploit the structural properties of ECG visibility graphs. Specifically, we integrate GATs and GINS to leverage their complementary strengths in capturing both structural and relational patterns within the graph representation of ECG signals.

The input to our model consists of a set of graphs $\mathcal{G} = \{G_1, G_2, ..., G_B\}$, where each graph $G_i = (V, E)$ comprises N nodes and E edges. Each node is initially represented by a feature vector of dimension 28, corresponding to signal characteristics extracted from the ECG waveform.

The architecture is structured as two GAT-GIN blocks, where each block has a GAT layer followed by a GIN layer. To enhance stability and expressivity, each block incorporates residual connections and Layer Normalization. Specifically, the residual connections sum the outputs of the GAT and GIN layers within each block after applying dropout, followed by Layer Normalization to stabilize training and prevent gradient instability.

Following the graph convolutional layers, a Global Attention Pooling (GAP) mechanism aggregates node embeddings into a graph-level representation of size B (batch size). This mechanism is parameterized by a gated neural network that adaptively determines the importance of each node embedding in the final graph representation.

The final classification stage consists of two fully connected layers, mapping the pooled representation to a $B \times 6$ output matrix. The output activation function $\sigma(x)$ (sigmoid) produces independent probability scores for each class, enabling multi-label classification. As detailed later, here we are interested in six heart abnormalities, and hence we consider six classes. The overall architecture is illustrated in Figure 4, and the following subsections provide a detailed breakdown of each processing step.

Graph Attention Network (GAT) With GAT layers, we deploy an adaptive attention mechanism that dynamically adjusts the contribution of each neighboring node according to its contextual significance. This approach proves particularly advantageous in the analysis of ECG graphs, where specific segments of the waveform are diagnostically more informative. The attention coefficient between nodes i and j is computed as:

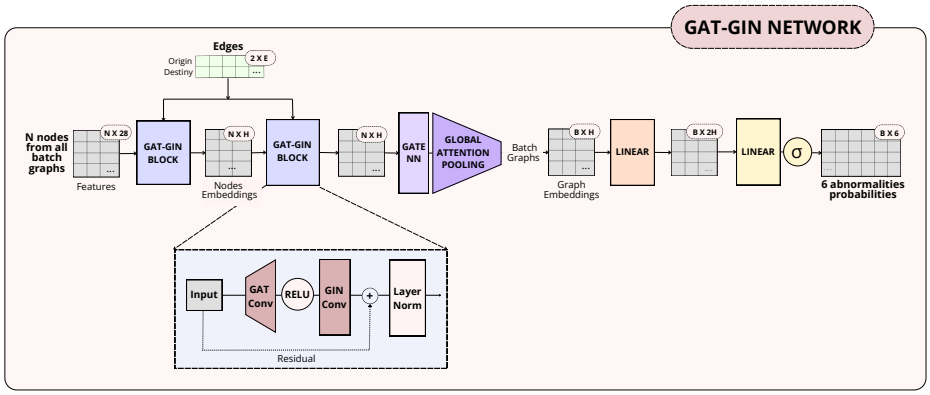

Fig. 4. Schematic of the proposed GNN-based architecture. The network comprises two GAT-GIN blocks, followed by a Global Attention Pooling layer and a fully connected classification module.

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T[Wh_i \parallel Wh_j]\right)\right)}{\sum_{k\in\mathcal{N}(i)} \exp\left(\text{LeakyReLU}\left(a^T[Wh_i \parallel Wh_k]\right)\right)}, \tag{1}$$

where Wh_i and Wh_j are the transformed feature vectors of nodes i and j, a is a learnable attention vector, and $\parallel$ denotes vector concatenation.

Graph Isomorphism Network (GIN) To complement GAT's attention mechanism, the GIN architecture was selected due to its strong discriminative power, crucial for differentiating between structurally similar yet clinically distinct ECG graphs. Given that cardiac abnormalities often induce subtle morphological changes in ECG waveforms, GIN's capacity to preserve graph structural information is particularly advantageous. Formally, the node embedding update in a GIN layer is defined as:

$$h_v^{(k)} = \text{MLP}^{(k)}\left((1+\epsilon^{(k)})\cdot h_v^{(k-1)} + \sum_{u\in\mathcal{N}(v)} h_u^{(k-1)}\right), \tag{2}$$

where $h_v^{(k)}$ represents the embedding of node v at the k-th layer, $\mathcal{N}(v)$ is the set of neighboring nodes of v, and $\epsilon^{(k)}$ is a learnable scalar parameter that adjusts the contribution of self-information.

Global Attention Pooling We employ a Global Attention Pooling (GAP) mechanism after our GNN layers, which dynamically determines the relevance of each node before aggregation. The GAP mechanism is parameterized by a gated neural network that adaptively weights each node representation, thereby preserving diagnostically significant information during aggregation:

$$h_G = \sum_{v \in V} \text{score}(h_v) \cdot h_v. \tag{3}$$

Fully Connected Classification Module The final stage of the model consists of a two-layer fully connected classifier, responsible for mapping the aggregated graph embedding to the six target classes. The transformation follows:

$$h' = \text{ReLU}(W_1 h_G + b_1), \tag{4}$$
$$y = \sigma(W_2 h' + b_2), \tag{5}$$

where W_1, W_2 are learnable weight matrices, and $\sigma(x)$ is the element-wise sigmoid activation function. The output is a $B \times 6$ matrix, where each row represents an independent probability estimate for the presence of a given abnormality.

This structured design allows the model to capture both local ECG waveform variations and global cardiac activity patterns, resulting in a robust classification framework.

3.3 Model Training and Optimization

Our model was trained for 50 epochs using the AdamW optimizer with an initial learning rate of 1×10^{-4} and weight decay of 5×10^{-4}. A hidden dimension of 64 was chosen to balance model capacity and computational efficiency. Dropout regularization with a rate of 0.4 was applied after each GNN and fully connected layer.

To improve generalization and stabilize the training process, the learning rate was halved every 10 epochs until reaching a lower bound of 5×10^{-5}.

The optimization criterion employed was the binary cross-entropy loss with logits, appropriate for multi-label classification:

$$\mathcal{L} = -\frac{1}{B} \sum_{i=1}^{B} \sum_{c=1}^{6} \left[y_{i,c} \log(\hat{y}_{i,c}) + (1 - y_{i,c}) \log(1 - \hat{y}_{i,c}) \right], \tag{6}$$

where $y_{i,c}$ is the ground truth label and $\hat{y}_{i,c}$ is the predicted probability for class c.

Thresholds for classifying ECG abnormalities were adjusted dynamically during validation. Each class had its decision threshold individually optimized to maximize the F1-score on the validation set.

For dataset splitting, we used the GroupShuffleSplit strategy, ensuring that exams from the same patient did not appear in both the training and validation sets. This method works by grouping data based on a common identifier (in this case, the patient) and then randomly splitting these groups into different sets while maintaining separation between individuals. This technique prevents data leakage and ensures a more realistic evaluation of the model's performance. The

model's performance was assessed using the macro-averaged F1-score, precision, recall, and confusion matrices.

Although training was performed for 50 epochs, we implemented an early stopping strategy based on the model's performance on the validation set, halting training when minimal improvements were observed in the validation metrics, ensuring better generalization.

4 Experiments and Results

4.1 Dataset

This study utilizes the CODE (Clinical Outcomes in Digital Electrocardiology) database [10], a large-scale ECG repository developed by the Telehealth Network of Minas Gerais (TNMG). Unlike artificially curated datasets, CODE reflects real-world clinical practice, preserving the natural noise, variability, and complexity of ECG recordings. This authenticity strengthens the external validity of machine learning models trained on the data, making them more applicable to real clinical scenarios.

Specifically, the complete CODE-15% dataset was employed, comprising 345,779 ECG exams from 233,770 unique patients, carefully curated to achieve balanced distributions across various age groups ranging from 16 to 85 years. This ensures demographic representativeness, reducing potential biases in classification performance. Each ECG exam contains recordings from the twelve standard leads (DI, DII, DIII, AVR, AVL, AVF, and V1–V6), offering a detailed depiction of the heart's electrical activity, and also covers six clinically significant ECG abnormalities: right bundle branch block (RBBB), left bundle branch block (LBBB), atrial fibrillation (AF), sinus tachycardia (ST), sinus bradycardia (SB), and first-degree atrioventricular block (1dAVb), selected for their distinct rhythm and morphological characteristics, making them particularly suitable for evaluating graph-based ECG classification methodologies.

From the total dataset, 312,406 ECG records were classified as unlabeled—free of any of the six investigated anomalies. The remaining 33,373 abnormal records are distributed as follows: 4,103 exams with 1dAVb, 7,566 with RBBB, 4,683 LBBB, 6,849 exhibiting ST, 4,830 with SB, and 5,432 diagnosed with AF. This large and realistic clinical distribution significantly contributes to the robustness and generalizability of our classification models.

Due to the high cost and complexity of graph-based data, we decided to reduce the amount of data used by selecting a subset of 115,260 ECGs for analysis (33% of CODE-15%). This reduction, while limiting the amount of available data, brings a significant advantage: it makes the implementation and use of the model more accessible for hospitals with limited resources, as well as simplifying training for healthcare professionals. By focusing the analysis on a smaller number of records, we were able to maintain the clinical relevance of the data while ensuring that the model remains practical and efficient for everyday clinical use, contributing to a faster and more effective deployment of machine learning-based technologies in medical practice.

4.2 Data Pre-processing

Before analysis, all ECG signals underwent rigorous pre-processing steps to enhance signal quality and reliability. Initially, records explicitly marked for interference or excessive noise within their clinical reports were excluded from the analysis and to further ensure signal quality, we applied frequency-based filtering to isolate clinically relevant cardiac information from potential interference sources [1]. Figure 6 demonstrates this filtering process, highlighting how the applied filters effectively mitigate unwanted noise and maintain essential ECG features.

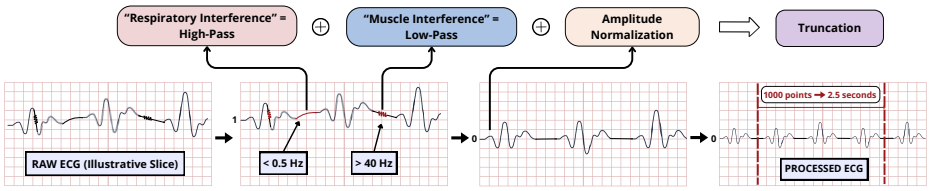

Fig. 5. Diagram illustrating the effect of the applied filters on the ECG signal. The high-pass filter removes respiratory interference (< 0.5 Hz), while the low-pass filter attenuates muscle interference (> 40 Hz), preserving the frequency components relevant for cardiac signal analysis.

First, we employed a high-pass filter with a cutoff frequency of 0.5 Hz to eliminate baseline wander, typically caused by patient movements and respiratory influences. Baseline wander can obscure critical waveform features by causing gradual drift in the ECG baseline. By permitting only frequencies above 0.5 Hz, we preserved essential characteristics of cardiac signals. Next, a low-pass filter with a cutoff frequency of 40 Hz was utilized to attenuate high-frequency noise arising from muscle artifacts and electrical interference, thus retaining frequency components most indicative of the heart's electrical activity. Combined, these filters create an effective band-pass filtering strategy, targeting the clinically significant frequency spectrum essential for accurate ECG interpretation.

Finally, a 2.5-second segment of the ECG signal, corresponding to 1000 sampling points, was isolated for analysis. Within this segment, R peaks—crucial landmarks for ECG interpretation—were identified. A window of 500 points before and after the central R peak was then extracted, ensuring that the complete waveform is captured for detailed analysis. If no R peak is detected, the midpoint of the segment is used as a reference. This extraction strategy standardizes the process and reinforces the reliability of subsequent clinical assessments of cardiac electrophysiology.

4.3 Experimental Setup

The experimental evaluation utilized our subset of the CODE-15% dataset for training and validation, which was partitioned into training (90%) and validation (10%) subsets. Given the limited amount of data, we performed class rebalancing to improve the model's training performance. As part of the balancing process, approximately 40% of the normal ECGs were removed to ensure a more equal distribution across the classes, enhancing the model's ability to learn from the minority classes and improving overall classification performance. The ECG records were randomly shuffled and stratified to maintain representative distributions of each abnormality class across all splits. To prevent data leakage, a patient-wise splitting strategy was employed, ensuring that no patient's exams appeared in multiple sets.

For testing, we used the CODE-test[12] dataset, which contains 827 ECG tracings from different patients and is annotated by multiple cardiologists, residents, and medical students. The dataset includes the same set of abnormalities as the CODE-15% dataset and has been reviewed by various professionals, making it a reliable and high-quality source for model evaluation.

All experiments were conducted on an NVIDIA RTX 3090 Ti GPU using PyTorch and PyTorch Geometric. ECG signals were transformed into VG using the ts2vg Python library, which employs a divide-and-conquer approach for efficient graph construction. This methodology allowed us to process large-scale ECG data while maintaining computational efficiency.

Performance evaluation was conducted using multiple metrics, including precision, recall, and F1-score. These metrics were calculated per class and averaged to provide a complete assessment of model performance. These hyperparameter settings were refined through iterative experimentation, analyzing loss curves and F1 scores to ensure optimal convergence and generalization.

4.4 Performance Evaluation

In this section, we present the performance of our proposed graph-based ECG classification model, comparing it with existing state-of-the-art deep learning baselines. We evaluate the model on the CODE-TEST dataset and report metrics including precision, recall, and F1-score for each of the six cardiac abnormalities.

Comparison with Baselines All baselines were trained on the complete CODE15% dataset, as described in their methodologies. Despite this, our approach outperforms all baselines in terms of F1-macro and recall when evaluated on the CODE-TEST dataset (Table 1), demonstrating its effectiveness even with fewer data. This suggests that the graph-based representation provides a strong alternative for capturing ECG patterns compared to sequential models. Having the highest recall is particularly important in medical applications, where missing a diagnosis can have serious consequences.

While our model does not surpass the ResNet architectures in precision, this is likely due to a stronger optimization for specificity at the cost of recall. Our

model, on the other hand, prioritizes capturing true positives, which is essential for detecting cardiac abnormalities reliably. These results evidences the trade-off between precision and recall in ECG classification and suggest that our approach is well-suited for clinical settings where minimizing false negatives is critical.

Table 1. Macro performance comparison (Precision, Recall and F1-score) for abnormality detection between GAT-GIN and other baseline models.

Metrics	ResNet-1 [11]	ResNet-2 [9]	ECG-Transform [4]	ECG-DETR [7]	GAT-GIN
Precision	0.875	**0.908**	0.711	0.777	0.803
Recall	0.778	0.743	0.687	0.661	**0.840**
F_1-score	0.814	0.811	0.677	0.699	**0.815**

Performance Comparison Across Abnormalities Table 2 presents the specific per-class F1-score for each of the six ECG abnormalities, along with the average F1-score across all classes. The results demonstrate the effectiveness of our approach, particularly when it comes to capture morphological patterns.

The GAT-GIN model set new records in detecting conduction abnormalities like 1dAVB and RBBB, where visibility graphs effectively preserved key morphological changes such as prolonged PR intervals and QRS complex alterations. However, CNNs showed slight advantages in LBBB and ST detection, likely due to their strength in capturing sequential features like QRS widening and rapid heart rate patterns.

Challenges emerged in classifying rhythm-based conditions like AF and SB, where diagnostic cues rely more on timing and variability than morphology. Irregular RR intervals in AF and subtle SB differences were harder for the graph-based model to distinguish. Future improvements could refine graph connectivity encoding or explore hybrid models that integrate sequential and structural learning for more comprehensive ECG classification.

Table 2. F1-score comparison between GAT-GIN and other baseline models for ECG abnormality classification.

Abnormality	ResNet-1	ResNet-2	ECG-Transform	ECG-DETR	GAT-GIN
1dAVB	0.661	0.719	0.489	0.631	**0.750**
RBBB	0.924	0.890	0.909	0.747	**0.927**
LBBB	**0.927**	0.843	0.886	0.826	0.909
SB	0.767	**0.821**	0.535	0.588	0.8
ST	**0.897**	0.833	0.763	0.838	0.857
AF	0.703	**0.758**	0.478	0.563	0.645
Avg.	0.814	0.811	0.677	0.699	**0.815**

4.5 Discussion

Our study presents a novel architecture for graph-based neural networks, addressing unexplored yet promising time series representations. Utilizing visibility graphs and GNNs, we achieved an F1 macro score of 0.815. The model outperforms competitive baselines, including ResNet-1, ResNet-2, ECG-Transform, and ECG-DETR, especially excelling in detecting RBBB and 1dAVB, while also showing strong performance for LBBB, SB, ST.

Due to the model's ability in capturing subtle morphological changes, the study outperformed the baselines despite using fewer data. In RBBB, for instance, signals changes may include an altered QRS complex, which would be reflected in the graph structure and the nodes relationships. The integration of GINs and GATs aim for these medical relevant dependencies, ensuring precise classifications.

Perhaps the most innovative perspective of our approach is its potential for interpretability. Figure 6 illustrates how visibility graphs for healthy and RBBB ECGs distinctly show morphological differences, which our model aims to capture. Moreover, this ability allows clinicians to visually trace decision-making processes and gain deeper insights into the rationale behind classifications, addressing the opacity of traditional deep learning models. Ultimately, the GAT layers dynamically adjust the relevance of different ECG segments, providing a transparent decision process that aligns with clinical markers.

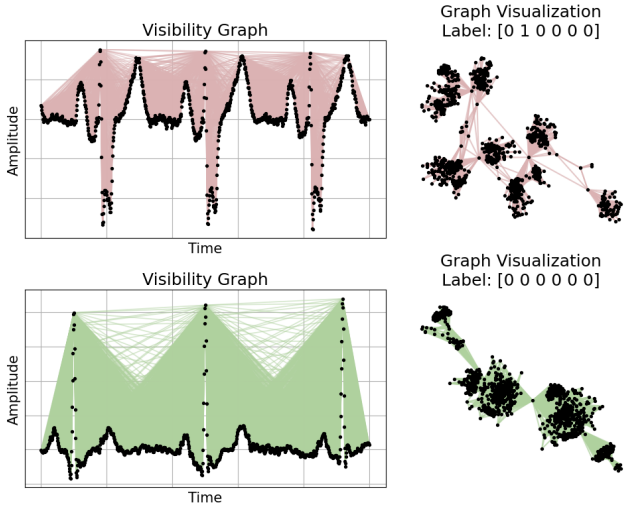

Fig. 6. The top signal (in red) corresponds to a patient with Right Bundle Branch Block (RBBB), a abnormality characterized by delayed electrical conduction in the right ventricle. The bottom ECG signal (in green) represents an unlabeled case.

5 Conclusion

5.1 Impact of Graph-Based Representations on ECG Classification

In this work, we introduced a novel method for ECG classification by transforming ECG signals into graph-based representations using visibility graphs and GNN's. This approach departs from traditional time-series analysis by focusing on the morphological and topographical features of the ECG signals. Our model achieved an F1 macro score of 0.815, outperforming conventional methods for conditions like RBBB and 1dAVB. The success of our approach demonstrates that visibility graphs not only improve classification accuracy, but also open new possibilities for clinical interpretability, enabling clinicians to gain deeper insights into the underlying ECG features that are critical for diagnosis.

5.2 Limitations and Future Improvements

Despite the promising results, our method faces practical challenges. Graph structures, particularly those derived from visibility graphs, are computationally heavy and require significant memory resources. This becomes a critical issue when processing large datasets, as both the conversion of ECG signals into graphs and the subsequent training of GNNs demand extensive VRAM and prolonged computational time.

Looking ahead, we propose several avenues for future improvement. One promising direction is to invert the ECG signals before generating the visibility graphs. This inversion may reveal additional features in the valleys, complementing the peak-focused information and potentially enriching the graph structure. Additionally, introducing weighted edges based on the distances between nodes could provide a more context-aware representation of the signal, thereby capturing subtler temporal dynamics. Finally and most importantly, we plan to explore the inherent interpretability and explainability offered by graph-based models. The explicit structure of graphs and the attention mechanisms in GNNs can be leveraged to trace back the decision process, offering valuable insights into the features driving the classification. This not only supports model validation, but also aligns with the increasing demand for transparent and clinically interpretable machine learning systems.

References

1. Berkaya, S.K., Uysal, A.K., Gunal, E.S., Ergin, S., Gunal, S., Gulmezoglu, M.B.: A survey on ecg analysis. Biomedical Signal Processing and Control **43**, 216–235 (2018)
2. Cheng, X., Wang, Z., Liu, K.: Time2graph+: Enhanced temporal feature learning via graph neural networks. Pattern Recognition **137**, 109276 (2023)
3. Coelho, J.V., Silva, M.T., Rodrigues, A.F.: Graph neural alchemist: A modular approach to time series representation via directed visibility graphs. arXiv preprint arXiv:2410.09307 (2024)

4. El-Ghaish, H., Eldele, E.: Ecgtransform: Empowering adaptive ecg arrhythmia classification framework with bidirectional transformer. Biomedical Signal Processing and Control **89**, 105714 (2024)
5. Ezz, M.: Deep learning-driven single-lead ecg classification: A rapid approach for comprehensive cardiac diagnostics. Diagnostics **15**(3) (2025)
6. Gopakumar, G., Nair, M., Xu, X., Wan, C., Anpalagan, A.: A deep learning method for ecg classification based on stacked autoencoders and cnn. Wireless Communications and Mobile Computing **2018**, 1–9 (2018)
7. Hu, R., Chen, J., Zhou, L.: A transformer-based deep neural network for arrhythmia detection using continuous ecg signals. Computers in Biology and Medicine **144**, 105325 (2022)
8. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.: From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences **105**(13), 4972–4975 (2008). https://doi.org/10.1073/pnas.0709247105
9. Rajpurkar, P., Hannun, A., Haghpanahi, M., Bourn, C., Ng, A.: Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint (2017)
10. Ribeiro, A., Paixão, G., Gomes, P., Ribeiro, M., Ribeiro, A., Canazart, J., Oliveira, D., Ferreira, M., Lima, E., Moraes, J., Castro, N., Ribeiro, L., Macfarlane, P.: Tele-electrocardiography and bigdata: The code (clinical outcomes in digital electrocardiography) study. Journal of Electrocardiology **57**, S75–S78 (2019)
11. Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P., Andersson, C.R., Macfarlane, P.W., Meira Jr, W., et al.: Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications **11**(1), 1760 (2020)
12. Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P., Andersson, C.R., Macfarlane, P.W., Meira Jr., W., Schön, T.B., Ribeiro, A.L.P.: Code-test: An annotated 12-lead ecg dataset (2020). https://doi.org/10.5281/ZENODO.3765780, `https://zenodo.org/record/3765780`
13. Subasi, A., Dogan, S., Tuncer, T.: A novel automated tower graph based ecg signal classification method with hexadecimal local adaptive binary pattern and deep learning. Journal of Ambient Intelligence and Humanized Computing **14**(2), 711–725 (Feb 2023). https://doi.org/10.1007/s12652-021-03324-4
14. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018), `https://openreview.net/forum?id=rJXMpikCZ`
15. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C.: Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). pp. 753–763 (2019)
16. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019)
17. Yang, S., Lian, C., Zeng, Z., Xu, B., Zang, J., Zhang, Z.: A multi-view multi-scale neural network for multi-label ecg classification. IEEE Transactions on Emerging Topics in Computational Intelligence **7**(3), 648–660 (2023)
18. Yao, Q., Wang, R., Fan, X., Liu, J., Li, Y.: Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network. Information Fusion **53**, 174–182 (2020)
19. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M.: Temporal graph convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems **21**(9), 3848–3858 (2020)