Graph-Based Temporal Learning on ECG Signals for Heart Condition Classification

Masoud Nateghi

***Abstract*—We propose a graph-based temporal learning framework for classifying heart conditions using 15-lead electrocardiogram (ECG) signals. By modeling ECG leads as nodes in a graph and applying Graph Neural Networks (GNNs) and Graph Attention Networks (GATs), our method captures both spatial dependencies among leads and temporal dynamics within each signal. We evaluate our approach on the PTB dataset, which contains 549 annotated ECG recordings labeled as either myocardial infarction or healthy control. Experimental results demonstrate that our convolutional GNN (CGNN) achieves strong classification performance with an F1-score of 73.85%, outperforming its attention-based counterpart. These findings highlight the effectiveness of temporal-aware graph learning for automated cardiac diagnosis.**

I. Introduction

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for approximately 20 million deaths worldwide in 2021 [1]. Electrocardiogram (ECG) is a fundamental, non-invasive diagnostic tool that records the electrical activity of the heart and provides critical information about cardiac function. Accurate and automated analysis of ECG signals can significantly improve diagnostic efficiency and potentially save lives through early detection[2].

While traditional machine learning and deep learning approaches have shown promising results in ECG analysis, they often fail to fully capture the complex relationships present in multi-lead ECG recordings. Standard CNNs treat ECG leads independently, ignoring the spatial relationships between leads. RNNs focus on temporal dependencies but struggle with the concurrent spatial relationships. Current approaches often lose valuable information in the transformation from raw signals to feature vectors.

Our graph-based approach addresses these limitations by explicitly modeling the relationships between ECG leads as a graph structure. It preserves both spatial and temporal information in a unified framework, potentially improving diagnostic accuracy for complex heart conditions that manifest across multiple leads with specific temporal patterns.

The authors are with the Department of Biomedical Informatics, Emory University (email: masoud.nateghi@emory.edu).

II. Related Work

Electrocardiogram (ECG) analysis has long been a focus of machine learning and signal processing research. Traditional methods for automated ECG classification include handcrafted feature extraction followed by shallow classifiers such as support vector machines (SVMs) and decision trees [3]. While effective in certain settings, these approaches are limited in their ability to generalize across large-scale, heterogeneous ECG datasets.

Deep learning models, particularly convolutional neural networks (CNNs), have shown strong performance in recent years by learning hierarchical representations of ECG signals directly from raw data [4, 5]. However, most CNN-based methods treat each lead independently, thereby ignoring the spatial relationships between different leads. Recurrent neural networks (RNNs), including LSTM and GRU variants, have also been explored to capture temporal dependencies in ECG sequences [6], but they similarly fall short in modeling inter-lead interactions.

Graph Neural Networks (GNNs) have recently emerged as powerful tools for modeling structured biomedical data, including biosignals. By treating multi-lead ECG recordings as graphs where each lead is a node, GNNs can naturally capture spatial dependencies between leads [7]. Attention-based GNNs, such as Graph Attention Networks (GATs), further enhance this modeling by enabling dynamic weighting of inter-lead relationships [8].

Our work builds upon these recent advances by introducing a temporal-aware graph-based pipeline for myocardial infarction detection using the PTB dataset. Unlike prior methods that often focus solely on spatial or temporal features, our approach integrates both aspects through a 1D-CNN-based feature extractor followed by GCN and GAT layers. We also conduct a direct comparison of convolutional and attention-based graph architectures, offering insight into their generalization behavior under data scarcity.

III. Dataset

We used the PTB (Physikalisch-Technische Bundesanstalt) dataset [9], which contains 549 high-resolution

ECG recordings from 290 subjects. These are longer recordings with variable duration, sampled at 1000 Hz, collected using a conventional 15-lead system. For our study, we focus on utilizing this dataset specifically for myocardial infarction detection, treating it as a binary classification task that distinguishes between patients with myocardial infarction and those without. The dataset includes 12 conventional leads plus 3 Frank leads (XYZ), providing comprehensive electrophysiological information critical for accurate detection of infarction patterns.

Among the 549 subjects, 384 recordings correspond to individuals diagnosed with myocardial infarction, while 165 recordings belong to patients with other conditions or healthy controls. This class imbalance is important to consider when evaluating model performance, especially in terms of recall and precision.

IV. Data Preprocessing

To ensure high-quality ECG signals for further analysis, a multistage preprocessing pipeline was implemented. The ECG preprocessing routine consists of the following steps: high-pass filtering to remove low-frequency noise, antialiasing filtering prior to downsampling, signal resampling to match the desired sampling rate, and notch filtering to eliminate power line interference.

A. Baseline Wander Removal

The first step addresses low-frequency baseline drift that is commonly caused by respiration or movement artifacts. A zero-phase low-pass filter is applied to isolate the baseline, which is then subtracted from the original signal. This subtraction acts as a high-pass filter, preserving the relevant ECG components while removing slow fluctuations.

B. Anti-aliasing filter and resampling

Prior to resampling, an antialiasing filter is applied to suppress high-frequency components that could alias into the signal's spectrum during downsampling. A Butterworth low-pass filter is used for this purpose and implemented in second-order sections (SOS) to ensure numerical stability.

The original ECG signals were sampled at 1000 Hz, a rate that captures a high degree of temporal detail. However, for many ECG analysis tasks (e.g., fiducial point detection, heart rate variability), such a high rate is not necessary and can be computationally inefficient. Therefore, the signals were downsampled to 360 Hz, which is widely recognized as sufficient to accurately represent the morphological features of ECG waveforms, including QRS complexes and P/T waves. This reduction lowers computational load and storage requirements without compromising clinical or analytical value.

C. Notch Filtering

A notch filter centered at 50 Hz is then applied to remove power line interference. Designed with a specified quality factor, this filter selectively attenuates the narrowband noise while preserving nearby signal components. Like the antialiasing filter, it is implemented in SOS form for robust performance.

V. Pipeline

A. Graph Representation of ECG Signals

We represent ECG signals as graphs, where each lead in a 15-lead ECG corresponds to a node (15 nodes in total). For each lead, we extract features using a 1D convolutional neural network (CNN) block consisting of two 1D convolutional layers with ReLU activations, followed by a pooling layer, as illustrated in Figure 1. This feature extractor transforms each lead's time-series signal into a fixed-size feature vector with 32 features, regardless of the original signal length. The resulting node embeddings for each subject form a matrix in $\mathbb{R}^{15\times 32}$.

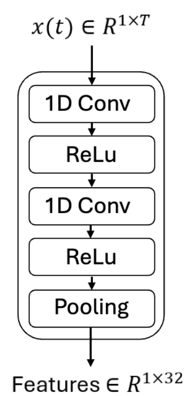

Fig. 1. Architecture of the feature extractor block applied to each ECG lead. The block consists of two consecutive 1D convolutional layers. The first convolutional layer uses 16 filters with a kernel size of 7, stride 2, and padding 3, followed by a ReLU activation. The second convolutional layer uses 32 filters with a kernel size of 5, stride 2, and padding 2, followed again by a ReLU activation. Finally, an adaptive average pooling layer reduces the temporal dimension to 1, resulting in a 32-dimensional feature vector for each lead.

B. CNN Based Graph Neural Network (CGNN)

Given the node embeddings of size $\mathbb{R}^{15\times 32}$ obtained from the feature extractor (described in Figure 1), we construct a graph where each node represents a lead with its corresponding 32-dimensional feature vector. The overall processing pipeline, illustrated in Figure 2, then applies two consecutive Graph Convolutional Network (GCN) layers, each followed by a ReLU activation function. Finally, the learned node features are passed through a multilayer perceptron (MLP) to perform binary classification into Myocardial Infarction (MI) or Healthy Control.

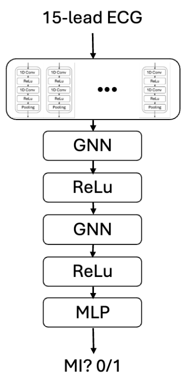

Fig. 2. The extracted node embeddings are passed through two consecutive GCN layers, each with 64 hidden units, and followed by ReLU activations. Global mean pooling aggregates node features across the graph, and a fully connected layer outputs the final classification (Myocardial Infarction vs. Healthy Control).

C. Graph Attention Neural Network (AttCGNN)

The AttCGNN architecture mirrors the CGNN pipeline, with the only difference being the replacement of Graph Convolutional Network (GCN) layers by Graph Attention Network (GAT) layers. Instead of uniformly aggregating neighbor features, GAT layers introduce learnable attention mechanisms to dynamically weight the contributions of neighboring nodes. This allows the model to emphasize the most informative inter-lead connections during feature learning. As in CGNN, the node features are processed through two consecutive GAT layers with ReLU activations, followed by a multilayer perceptron (MLP) for final classification.

VI. Results

We evaluate the performance of two graph-based neural network architectures for myocardial infarction detection using the PTB dataset. Both models were trained for 200 epochs, and performance was assessed using training and validation accuracy and loss metrics.

Figure 3 and Figure 4 illustrate the training dynamics of the CGNN model. The training loss steadily decreases over time, indicating successful minimization of classification error. While the validation loss exhibits fluctuations, it remains relatively stable throughout training. In terms of accuracy, the CGNN achieves a consistent increase in training accuracy, surpassing 90% by the end of training. Validation accuracy also improves gradually, reaching a peak of approximately 84%, suggesting good generalization to unseen data. These results demonstrate that the CGNN is capable of effectively extracting relevant features from multi-lead ECG signals.

In contrast, the attention-based attCGNN model, presented in Figures 5 and 6, shows a divergent training behavior. The training accuracy rapidly increases, eventually approaching 100%, and the training loss converges toward zero. However, the validation loss displays substantial fluctuations and an overall upward trend, with values exceeding 2.0 in later epochs. The validation accuracy remains low and unstable, oscillating around 70–75% without a clear upward trajectory. This stark discrepancy between training and validation performance indicates severe overfitting in the attCGNN model.

In addition to accuracy and loss, we evaluated both CGNN and attCGNN models using precision, recall, and F1-score to better assess clinical reliability, particularly in the context of imbalanced classes in myocardial infarction detection. Table I summarizes the classification performance on the validation set for both models.

TABLE I
VALIDATION PERFORMANCE METRICS FOR CGNN AND ATTCGNN MODELS.

Model	Accuracy	Precision	Recall	F1-score
CGNN	0.8455	0.7742	0.7059	0.7385
attCGNN	0.80	0.7308	0.5588	0.6333

As shown in Table I, the CGNN achieves the highest F1-score of 0.7385, indicating a good balance between precision and recall. Its precision (0.7742) suggests that false positives are relatively well-controlled, while a recall of 0.7059 implies a strong ability to detect true myocardial infarction cases.

In contrast, the attCGNN model, although achieving high training performance, underperforms across all metrics on the validation set. Its lower F1-score of 0.6333 reflects poor generalization, likely due to overfitting. This is consistent with the previously discussed accuracy and loss trends.

These results underscore the CGNN model's robustness and potential clinical utility, while highlighting the need for further optimization of attention-based models in data-scarce settings. Overall, while both models learn effectively on the training set, the CGNN outperforms the attention-based variant in terms of validation performance. The attention mechanism, though powerful in theory, may require additional regularization or architectural tuning to be effective in this context. These findings highlight the importance of balancing model capacity and generalization when applying deep learning methods to clinical ECG data.

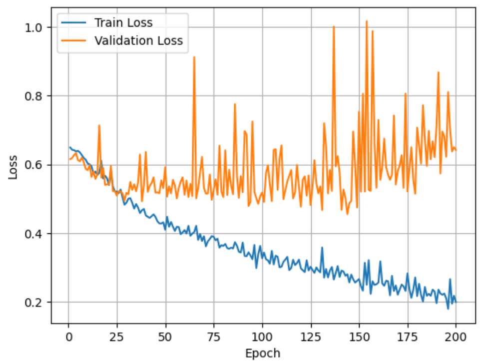

Fig. 3. Loss values for the Convolutional Graph Neural Network (CGNN) model over 200 epochs of training on the PTB dataset for myocardial infarction detection. The blue line represents training loss, which shows consistent decrease throughout training, while the orange line represents validation loss.

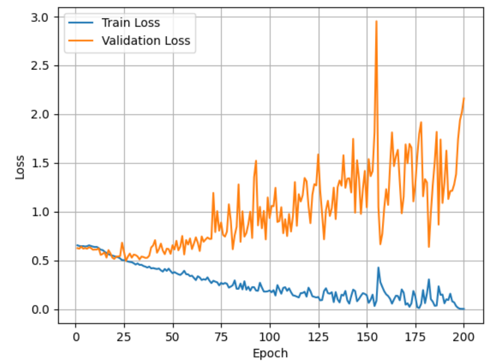

Fig. 5. Training and validation loss curves for the attention-based graph neural network (attCGNN) over 200 epochs. The training loss consistently decreases, while the validation loss shows high variance and an overall increasing trend, indicating potential overfitting.

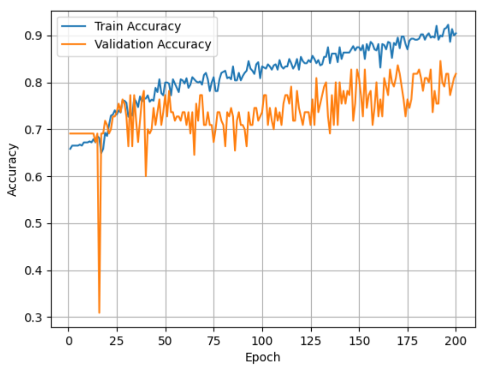

Fig. 4. Accuracy metrics for the Convolutional Graph Neural Network (CGNN) model over 200 epochs of training on the PTB dataset for myocardial infarction detection. The blue line represents training accuracy, which steadily increases to over 90%, while the orange line shows validation accuracy reaching a peak of approximately 84%.

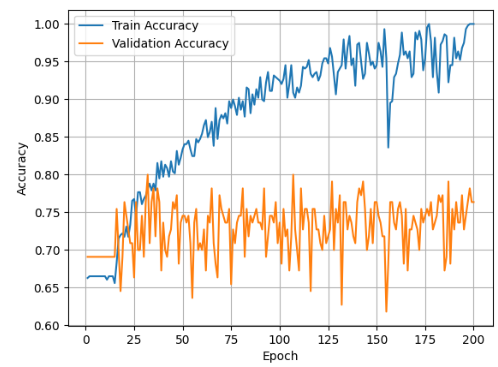

Fig. 6. Training and validation accuracy curves for the attention-based graph neural network (attCGNN) over 200 epochs. While training accuracy approaches 100%, validation accuracy remains low and unstable, reflecting poor generalization and overfitting.

VII. Discussion

This study investigates the effectiveness of GNN architectures for myocardial infarction (MI) detection using the PTB ECG dataset. Specifically, we compare a Convolutional Graph Neural Network (CGNN) and an attention-based GNN (attCGNN), where ECG leads are modeled as graph nodes and features are extracted using convolutional mechanisms.

The CGNN achieved the best overall performance, with a validation accuracy of 84.55%, a precision of 77.42%, a recall of 70.59%, and an F1-score of 73.85%. These results indicate strong generalization capability and a relatively balanced trade-off between sensitivity (recall) and specificity (precision). The attCGNN model lagged in performance with a validation accuracy of 80.00% and a significantly lower recall of 55.88%, resulting in an F1-score of 63.33%. This drop in recall suggests that the attention-based model misses more true positives than the CGNN, which could be critical in a diagnostic setting where sensitivity is paramount.

The superior performance of CGNN suggests that it effectively captures the spatial dependencies between ECG leads, offering a stable learning process with less overfitting. In contrast, the attCGNN's lower recall and F1-score hint at potential instability or overparameterization relative to the dataset size. Despite attention mechanisms being theoretically suited to modeling complex relationships, they may require larger datasets or more refined tuning to outperform simpler architectures like CGNN in practice.

VIII. Conclusion

We proposed a graph-based approach for myocardial infarction detection from ECG signals by modeling individual leads as nodes and extracting features using convolutional layers. Two GNN-based architectures were developed: CGNN and attCGNN. Experimental results on the PTB diagnostic dataset show that the CGNN outperforms the attention-based variant across all key evaluation metrics, including accuracy, precision, recall, and F1-score.

The CGNN achieved a validation accuracy of 84.55% and an F1-score of 73.85%, demonstrating its ability to generalize well on unseen data. In contrast, attCGNN, while achieving an accuracy of 80.00%, showed a lower F1-score of 63.33%, mainly due to its reduced recall. These findings indicate that convolution-based message passing is currently more reliable than attention-based methods for ECG graph classification tasks under limited data conditions.

Our study reinforces the value of graph neural networks for clinical ECG analysis and shows that explicit modeling of inter-lead relationships can enhance diagnostic performance. These contributions add to the growing body of evidence supporting the integration of GNNs into medical signal processing pipelines.

Supplementary Materials

The code and data for reproducing the results in this paper can be found and downloaded at the link: https://github.com/MasoudNateghi/CS-584.

References

[1] M. Di Cesare, P. Perel, S. Taylor, C. Kabudula, H. Bixby, T. A. Gaziano, D. V. McGhie, J. Mwangi, B. Pervan, J. Narula, D. Pineiro, and F. J. Pinto, "The heart of the world," *Global Heart*, vol. 19, no. 1, 2024.

[2] M. Martínez-Sellés and M. Marina-Breysse, "Current and future use of artificial intelligence in electrocardiography," *Journal of Cardiovascular Development and Disease*, vol. 10, p. 175, Apr. 2023.

[3] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.

[4] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[5] P. Rajpurkar, A. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.

[6] O. Yildirim, U. B. Baloglu, R. S. Tan, E. J. Ciaccio, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ecg signals," *Computers in biology and medicine*, vol. 102, pp. 411–420, 2018.

[7] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, 2021.

[8] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2018.

[9] R.-D. Bousseljot, D. Kreiseler, and A. Schnabel, "The PTB Diagnostic ECG Database," 2004.