



Fusion of edge detection and graph neural networks to classifying electrocardiogram signals

Linh T. Duong^{a,b,1}, Thu T.H. Doan^{c,1}, Cong Q. Chu^d, Phuong T. Nguyen^{e,*}

^a National Institute of Nutrition, Vietnam Ministry of Health, Viet Nam

^b Institute of Research and Development, Duy Tan University, Viet Nam

^c University of Engineering & Technology, Vietnam National University, Viet Nam

^d Hanoi Oncology Hospital, Viet Nam

^e Department of Information Engineering, Computer Science and Mathematics University of L'Aquila, 67100, L'Aquila, Italy

ARTICLE INFO

Keywords:

Deep learning
Graph neural network(s)
Electrocardiogram (ECG/EKG)
Bio-signaling
Healthcare

ABSTRACT

The analysis of electrocardiogram (ECG) signals are among the key factors in the diagnosis of cardiovascular diseases (CVDs). However, automatic processing of ECG in clinical practice is still restrained by the accuracy of existing algorithms. Deep learning methods have recently achieved striking success in a variety of task including predictive healthcare. Graph neural networks are a class of machine learning algorithms which can learn by directly extracting important information from graph-structured data, and perform prediction on unknown data. Such algorithms are suitable for mining complex graph data, deducing useful predictions. In this work, we present a Graph Neural Network (GNN) model trained in two datasets with more than 107,000 single-lead signal images extracted from laboratories of Boston's Beth Israel Hospital and of the Massachusetts Institute of Technology (MITBIH), and 1.5 million labeled exams analyzed by the Physikalisch-Technische Bundesanstalt (PTB). Our proposed GNN achieves promising performance, i.e., the results show that ECG classification based on GNNs using either single-lead or 12-lead setup is closer to the human-level in standard clinical practice. By several testing instances, the proposed approach obtains an accuracy of 1.0, thereby outperforming various state-of-the-art baselines by both databases with respect to effectiveness and timing efficiency. We anticipate that the approach can be deployed as a non-invasive pre-screening tool to assist doctors in real-time monitoring and performing their diagnosis activities.

1. Introduction

Several studies indicate that Cardiovascular Diseases (CVDs) are the leading cause of human death worldwide (Benjamin et al., 2017; Smith & Jones, 2012). In fact, CVDs account for 31% of the deaths in 2016, from which 85% happened due to heart attack. According to a practical scenario, cardiologists usually give an early diagnosis based on the electrocardiogram (ECG) representing for electrical signals in our heart (Darmawahyuni et al., 2022). ECG is a non-stationary physiological signal, and it is one of the most commonly used non-invasive diagnostic tools, which might aid doctors in making a reasonable initial diagnosis. Automated ECG analysis has become obviously trendy in modern medicine. It brings numerous advantages which could be considered as reducing health care costs, minimizing medical errors, optimizing clinical workflow, and facilitating clinical diagnosis. Given

the circumstance, a quick and accurate diagnosis without human intervention would be obviously crucial to the management and treatment of heart diseases.

Machine learning (ML) techniques have recently become more widespread, enabling a wide range of applications across various domains (Portugal, Alencar, & Cowan, 2018). A numerous number of studies have been dedicated to apply ML algorithms to solve different issues in the healthcare sector (Duong, Nguyen, Iovino and Flammini, 2023; Soltanisehat, Alizadeh, Hao, & Choo, 2020). Some early approaches have been proposed by applying deep learning (DL) techniques on ECG images to automatically detect heart problems (Rasti-Meymandi & Ghaffari, 2022). In fact, to have a good enough model based on these techniques, it is necessary to train them with large patient datasets, which are not always available in healthcare systems. Otherwise, it is highly probable that many existing techniques gain limited accuracy

* Corresponding author.

E-mail addresses: duongtuanlinh@dinhuong.org.vn, duongtuanlinh@duytan.edu.vn (L.T. Duong), doanthihoaitu@vnu.edu.vn (T.T.H. Doan), congq@bvubhn.vn (C.Q. Chu), phuong.nguyen@univaq.it (P.T. Nguyen).

¹ The first two authors contributed equally to the paper.

Table 1
A summary of the related studies.

Approach	Dataset	Size	# classes	Techniques	AUC	Acc. ^a
Hannun et al. (2019)	ECG	53,549 pat. ^b	12	CNN	0.970	N/A
Attia et al. (2019)	ECG	97,829 pat.	2	CNN	0.930	0.857
Noseworthy et al. (2020)	ECG	97,829 pat.	12	CNN	0.930	N/A
Abdalla et al. (2019)	ECG	10,782 rec. ^c	5	ANN	1.00	0.999
Fu et al. (2021)	ECG	5479 rec.	20	CNN & RNN	0.901	0.816
Zhu et al. (2020)	ECG	70,692 pat.	9	CNN	0.983	N/A
Kashou et al. (2020)	ECG	720,978 pat.	66	ResNet	0.980	N/A
Tison et al. (2018)	ECG	9750 part. ^d	2	CNN	0.970	N/A
Zhao, Liu, Han, and Peng (2022)	ECG	21,837 rec. ^c	5	GNN	0.937	0.9
		6877 rec. ^c	9			

N/A corresponds to not applicable.

^aDenotes Accuracy.

^bIndicate patients.

^cIndicate recordings.

^dIndicate participants.

and require long time on training/testing process, making them unsuitable for real-time usage. Furthermore, these approaches could be considered as “black-boxes” as it is not clear how they make their decision. That means it would be very challenging to consistently increase performance on specific tasks.

Neural networks are among the most notable ML techniques, which learn from labeled data to deal with unlabeled data afterwards (Svozil, Kvasnicka, & Pospíchal, 1997). Neural networks have been widely used on different tasks with various types of data (*text, image, audio*) (Bishop, 1995; Sellami & Hwang, 2019; Zhang, Eddy Patuwo, & Hu, 1998). Especially, the evolution of Convolutional Neural Networks (CNNs) contributes to a huge step forward in image and speech recognition (He, Zhang, Ren, & Sun, 2015). In this respect, CNNs have been exploited to classify heart diseases in ECG images, and they demonstrated to achieve encouraging prediction performance (Attia et al., 2019; Fu, Hong, Zhang, & Du, 2021; Kashou et al., 2020; Noseworthy et al., 2020; Zhu et al., 2020). However, getting a better accuracy comes at a price: CNNs need to be scaled up wider, deeper, and/or higher resolution, and this leads to new architectures that have more parameters, which raises the cost of training and testing to an unaffordable level. Therefore, various compact CNNs have been established in an effort to preserve the balance between accuracy and computing expense (Le, Lin, & Piedad, 2019). Still, depending on the purposes, we need to investigate more suitable techniques to further improve prediction performance.

In this paper, we propose a practical solution to improve heart disease prediction performance, exploiting the power of graph neural networks for ECG data. First, we convert all ECG recordings in these datasets into waveform images following their annotation. Second, we employ the Sobel operator (Kanopoulos, Vasanthavada, & Baker, 1988) for edge detection, transforming the waveform images into a graph-structured format. The processed data is then used to feed a GNN model. Our approach has been evaluated on two datasets, namely MIT-BIH Arrhythmia (Goldberger et al., 2000) and PTB-XL (Wagner et al., 2020). A series of experiments have been executed on the proposed ECG-GNN using the aforementioned datasets, comparing it with various state-of-the-art approaches built on top of convolutional neural networks, including ResNet and ConvNet. The evaluation shows that our model yields promising prediction results both in terms of efficiency and effectiveness.

In this respect, our paper has the following contributions:

- A practical approach to detect edges of waveform in ECG images, using the Sobel operator.
- A workable solution to the classification of heart diseases, deploying GNN techniques on ECG signals. *To the best of our knowledge, our study is the first attempt to deploy GNNs in automatically classifying ECG signals for detecting heart problems.*
- An empirical evaluation on the two real ECG datasets to compare our proposed model with two state-of-the-art approaches using convolutional neural networks (ResNet26D and ConvNet).

The paper is organized as follows. To associate our work with state-of-the-art research, we review several relevant studies in Section 2. Section 3 provides background related to electrocardiogram and graph neural networks. Afterwards, our proposed approach to the classification of ECG is presented in Section 4. In Section 5, the datasets and metrics used for evaluation are described in detail. After that, Section 6 presents and analyzes our experimental results. Various threats to validity of our achieved results are also discussed in this section. Finally, Section 7 sketches a few up-and-coming work in the future, and concludes the paper.

2. Related work

So far, there have been various approaches using deep learning techniques to classify ECG signals to identify heart diseases. We consider the most notable studies in our review, paying attention to the dataset used and the final prediction accuracy. Table 1 provides a summary of the related work being considered in this section.

With the purpose of modeling the relationship between ECG 12-leads in ECG anomaly recognition, Zhao et al. proposed a novel model, called ECGNN (Zhao et al., 2022), which employs a graph neural network to learn lead-level representations and graph-level representations. This approach consists of a feature extractor backbone and a graph neural network module, with the former being a neural network to extract features of ECG signal for initialization of the graph nodes, and the latter combines graph convolution and graph pooling to generate hierarchical representations of graphs. They performed evaluation on the PTB-XL (Wagner et al., 2020) and ICBE 2018 datasets (Liu et al., 2018), and yielded 0.94 as the maximum accuracy. With respect to our proposed approach, we build our model by employing Sobel operator (Kanopoulos et al., 1988) and multiple GNN layers, which show more impressive performance in the same task.

Hannun et al. (2019) investigated the efficiency of a DNN model performed on cardiologist-level arrhythmia detection and classification. The dataset used in this work includes 91,232 single-lead ECG records, collected from 53,549 patients who used a single-lead ambulatory ECG monitoring device. The DNN was designed to classify 12 output rhythm classes, including 10 arrhythmias, sinus rhythm and noise. The authors reported an AUC of 0.97 as the highest. This work demonstrates that an end-to-end DNN model raises a significant improvement in ECG interpretation.

Recently, CNNs have shown to achieve promising performance in image processing applications (Duong, Le, Tran, Ngo, & Nguyen, 2021). In the healthcare domain, CNN is also performed on ECG images to tackle different tasks related to heart disease prediction. We review various studies using CNNs that are relevant to our current work.

Attia et al. (2019) developed a CNN model to classify 2-class ECG dataset of 97,829 patients. This study determines whether or not a patient get asymptomatic left ventricular dysfunction. They performed

training process on a dataset collected from 44,959 patients and tested with an independent set of 52,870 patients. Their model obtained 0.93 of AUC, and 0.857 of accuracy. Secondly, extended from the previous study by using the same dataset, [Noseworthy et al. \(2020\)](#) assess the impact of race and ethnicity on the performance of their designed CNN model. They aim to identify whether the derivation population or racial variation in the ECG drives the model performance. Their conclusion indicates that the variety by race in ECG characteristics does not impact the ability of the CNN model to predict low left ventricular ejection fraction. Additionally, to guarantee the responsibility of machine learning algorithms in medicine, the authors encourage more investigations on new AI technologies among diverse ethnic, racial, age, and sex groups.

A recent study proposes a CNN model ([Zhu et al., 2020](#)) for real-time ECG analysis. It aims to generate the diagnosis for arrhythmias at multilabel level. There are 21 distinct rhythms as separate classification labels, including most types of heart rhythm or conduction abnormalities. The model in this study achieves the best AUC of 0.983 on a dataset comprised 180 112 ECG records from 70,692 patients.

An other serious heart disease is Atrial fibrillation (AF) which is considered as the leading cause of stroke. A related study ([Tison et al., 2018](#)) presents a deep neural network model to detect AF automatically based on the reference standard of 12-lead ECG-diagnosed AF. The dataset in this research is produced by smartwatches from 9750 participants, including 347 participants with AF. The authors train their designed model by using a method called heuristic pretraining, in which the network is represented approximately by the R-R interval. The maximum AUC achieved in this study is 0.97. [Abdalla et al. \(2019\)](#) also work on detecting different types of arrhythmia from heartbeat information. Their proposed approach in this study obtains impressive results in an experiment with a small dataset of 10,782 ECG recordings. Their selected dataset is downloaded from Physionet in the MIT-BIH database, classifying five different arrhythmia heartbeats. They develop an artificial neural network (ANN) based on the non-linearity and nonstationary decomposition methods. Completing ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was used to obtain intrinsic mode functions (IMFs). Their proposed model yields an outstanding final result with values 1.00 for AUC and 0.99 for accuracy.

Another different approach combining both CNN and RNN techniques is presented in recent work ([Fu et al., 2021](#)). It is performed on a very small dataset, which consists of 5479 ECG recordings collected from human body by a small hardware device. In order to accomplish automatic and accurate cardiovascular disease diagnosis, a novel deep-learning-based cloud service was developed and deployed. The designed model in this study supports 20 types of diagnostic items including sinus rhythm, tachyarrhythmia, and bradyarrhythmia, 17 of them have more than 0.98 ROC-AUC score. Otherwise, the obtained AUC and accuracy are 0.901 and 0.816, respectively.

[Kashou et al. \(2020\)](#) proposed an algorithm using CNN as a multilabel classifier for categorizing various heart problems. Based on the cardiologist's final annotation, 66 discrete corresponding to different heart issues are transformed into diagnostic ECG codes. Their research is performed on a massive dataset including 2,499,522 ECGs from 720,978 patients. The authors applied ResNet model ([He et al., 2015](#)) as a classifying engine. The proposed approach achieved at least of 0.98 for AUC value.

3. Background

This section first provides background on preprocessing data phases and constructing Graph-structured data. Afterwards it gives a briefly introduction to an architecture of our proposed GNN model, which is used as the classification engine in our work.

3.1. Edge detection with the Sobel operator

The Sobel operator is considered as a simple but effective method for image edge detection ([Kanopoulos et al., 1988](#)). It is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. Technically, it multiplies an input image with convolution matrices to yield sharp discontinuities. In particular, edge detection is performed both in the horizontal and vertical directions, using the following two convolution matrices.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } G_y = \begin{bmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Essentially, the Sobel operator is a weighted average filter as explained as follows. At every point in an input image, the result of the operation is either the corresponding gradient vector, or the norm of this vector. The filter is the partial derivative of $f(x, y)$ as the central computing 3×3 neighborhood at the x, y direction. In order to eliminate noise, a certain weight is correspondingly increased on the centre point, and the digital gradient approximation equations are described as follows.

$$\begin{aligned} G_x &= \{f(x+1, y-1) + 2 * f(x+1, y) + f(x+1, y+1) \\ &\quad - \{f(x-1, y-1) + 2 * f(x-1, y) + f(x-1, y+1)\} \\ G_y &= \{f(x-1, y+1) + 2 * f(x, y+1) + f(x+1, y+1) \\ &\quad - \{f(x-1, y-1) + 2 * f(x, y-1) + f(x+1, y-1)\} \end{aligned}$$

The main advantage of the operator is its fast detection speed, as well as the ability to remove noise. Due to these characteristics, the filter is employed in this work to detect waveform in ECG images, attempting to increase the prediction performance.

3.2. Graph neural networks

Deep Learning techniques such as Convolutional Neural Networks, Recurrent Neural Networks, autoencoders, or transformers ([Duong et al., 2021](#)) are successful in dealing with hidden pattern of Euclidean geometry data, e.g., images, texts, or videos. However, there is another type of data called non-Euclidean geometry, and such data is represented as graphs with complex relationship and interdependencies between objects. For this type of data, the aforementioned deep learning techniques appear to be less applicable.

A graph is a structure to analyze the pair-wise relationship between objects and entities by means of two components, i.e., *vertices* and *edges*, $G = (V, E)$, where V is a set of *vertices* and E is the edges between *nodes*. If there is a directional dependency between nodes occur, then the edges are directed; otherwise, edges are undirected. Correspondingly, graph neural networks have been developed to deal with this type of data ([Wu et al., 2021](#)).

A graph is often represented as an adjacency matrix A of size $(N \times N)$, where N is the number of nodes. If each node is characterized by a set of M features, then a dimension of feature matrix X is $(N \times M)$. Graph-structured data is complex, and thus it brings a lot of challenges for existing machine learning algorithms. Hence, GNNs are designed to perform inference on data described by graphs. By simplifying the problems into simpler representations or transform them into representations from different perspectives, GNN models enable to solve complex problems. In fact, GNNs are a good way to deal with abstraction concepts such as relationships and interactions. As how GNNs are named, they contribute to a convenient method for edge-level, node-level, and graph-level classification task, graph visualization and graph clustering.

We call u and v as two nodes in a graph, and x_u and x_v their corresponding feature vectors. The encoder function $Enc(u)$ and $Enc(v)$ convert the feature vectors to z_u and z_v .

$$h_v^0 = X_v(\text{feature_vector}) \quad (1)$$

$$h_v^k = \sigma(W_k \sum \frac{h_v^{k-1}}{|N(v)|} + B_k h_v^{k-1}) \quad (2)$$

where $W_k \sum \frac{h_v^{k-1}}{|N(v)|}$ is the average of all the neighbors of node v , while $B_k h_v^{k-1}$ is the previous layer embedding of node v augmented with B_k – a trainable weight matrix and it represents a self-loop activation for node v .

In the k th encoder layer, given a target node i , the importance of a neighboring node $j \in \Gamma$ is computed as:

$$e_{(k)}^{ij} = \sigma(a_s^{(k)T} W^{(k)} h_i^{k-1} + a_r^{(k)T} W^{(k)} h_j^{k-1}) \quad (3)$$

where $W^{(k)}$ represents the trainable parameters of the k th encoder layer. The importance coefficients of a node comparable is normalized by the softmax function:

$$\alpha_{ij}^{(k)} = \frac{\exp(e_{ij}^{(k)})}{\sum_{l \in N_i} \exp(e_{il}^{(k)})} \quad (4)$$

where N_i represents the neighboring nodes of node i . Eq. (4) normalizes the sum of the importance scores of all neighboring nodes to 1. The final representation of node i output by the k th encoder layer is computed using the following formula:

$$h_i^{(k)} = \sum_{j \in N_i} \sigma(\alpha_{ij}^{(k)} W^{(k)} h_j^{(k-1)}) \quad (5)$$

By *link-level prediction*, it is necessary to represent the relationship among nodes in graph and predict if there is a connection between two entities. Meanwhile in *node-level classification*, the task is to understand node embedding for every node in a given graph by looking at the labels of their neighbors. With *graph-level classification*, the entire graph needs to be classified into suitable categories. This is similar to image classification, yet the target substitutes for the graph domain. Graph classification can be applied in a wide range of both academic and industrial problems, e.g., classifying a molecular structure into a meaningful category.

Using popular CNNs and/or Transformers, models can recognize objects in images and videos, obtaining promising accuracy (Duong, Chu, Nguyen, Nguyen and Tran, 2023; Garcia & Bruna, 2017; Marino, Salakhutdinov, & Gupta, 2017; Wang, Chen et al., 2018; Wang, Ye and Gupta, 2018). In the scope of this paper, we employ different GNN techniques following the graph classification paradigm to the detection of mammograms. The next section introduces our proposed approach in detail.

4. Proposed approach

This section presents our proposed approach to realize a system for automated classification of ECG signals. In particular, in Section 4.1, we explain how to transform the input images into the graph format using the Sobel algorithm (Kanopoulos et al., 1988). Afterwards, Section 4.2 introduces the classification engine built on top of graph neural networks.

4.1. Data transformation

Graph neural networks accept as input data prepared in a graph format with nodes and edges. We transform the original images in the DICOM format (Digital Imaging and Communications in Medicine) and converted them to lighter format as Joint Photographic Experts Group (JPEG), which are then rescaled to $512 \times 512 \times 1$.

ECG recordings are converted to waveform images following their annotation. To get better performance, we set all preprocessed images size of $64 \times 64 \times 1$. Afterwards, these transformed ECG signal images are used to constructed graph format data.

Following the convolution of each 3×3 sub-matrix by filters on both the horizontal and vertical axes, gradient for each sub-matrix is calculated. Because of the conversion of all grayscale images, each pixel can be altered if the gradient magnitude is larger than the gray value of 128 (Saha et al., 2021). The edge maps are then used to build proper graphs for each dataset. Afterward, the Sobel operator (Kanopoulos et al., 1988) is employed to detect edge of waveform images, which are then converted into graph data by following the steps below:

- A pixel with grayscale intensity value equal to or larger than 128 is converted to a node. A node feature contains the intensity of the relative pixel;
- An edge is used to connect vertexes, grouping neighboring pixels in the original images;
- A graph is constructed from a single image, nodes and edges are built from a relevant image the property of the same graph. Attributes of the vertex known as grayscale intensity values are normalized graph-wise. Normalization operator is performed by subtracting the mean of all attributes under every graph from the original value of preprocessed image and then dividing it by the standard deviation.

4.2. Classification engine

The proposed architecture consists of two main phases, i.e., training and testing as shown in Fig. 1. In the preprocessing phase, the first step is to convert all ECG recordings to waveform images following their annotation. Each ECG recording consists of a set of electrical signals which is stored in digital format. From their annotation file, the corresponding images are constructed as input for the classification model in the next steps. Afterward, the Sobel operator (see Section 3.1) is employed to filter out noise, and detect waveform in ECG images, allowing us to increase the prediction performance.

We used GraphConv layers (Li, Han, & Wu, 2018; NT & Maehara, 2019; Oono & Suzuki, 2021), which consists of a layer architecture originated from a theoretical point of view and it is related to the 1-dimensional Weisfeiler-Leman graph isomorphism heuristic (1-WL) (Leman & Weisfeiler, 1968). A generalization of 1-GNNs, so-called k-GNNs, has been developed based on the k-WL. Its layer can take higher-order graph structures at multiple stack into statement. We build a Graph Neural Network with only three layers, and three GNN models based on a backbone of GraphConv (Morris et al., 2018). Input and output channels at the first GNN layer are equal to the number of node features of relevant input data and 64. The next two input and output channels of the other GNN layers are set as 64. Following two first GNN layers, there are activation function of ReLu (Agarap, 2018) and Dropout (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) of 0.5, whereas the last GNN layer is connected to ReLu activation and a linear layer with input of 64 and output as numbers of categories.

The rationale behind the selection of three layers is as follows. Existing research has shown that Graph Neural Networks work better with scale ratios of narrow width and shallow depth. In particular, in a previous research (Loukas, 2020), the author conducted a series of experiments following the scalability of ratios in depth and width of GNN. These results show that GNNs' performance in both training and testing phases is reduced, given that either deeper or wider models are employed. In this paper, we empirically conducted a range of experiments such as filter operations, layer types of GNN, and ratios of width and depth as well.² Eventually, we realized that the best trade off between efficiency and effectiveness is obtained using three layers in depth combined with width channels equal to 64.

Table 2 displays information about the used backbones in terms of the number of parameters, the model size, and the training time in

² For the sake of clarity, we do not present these results here.

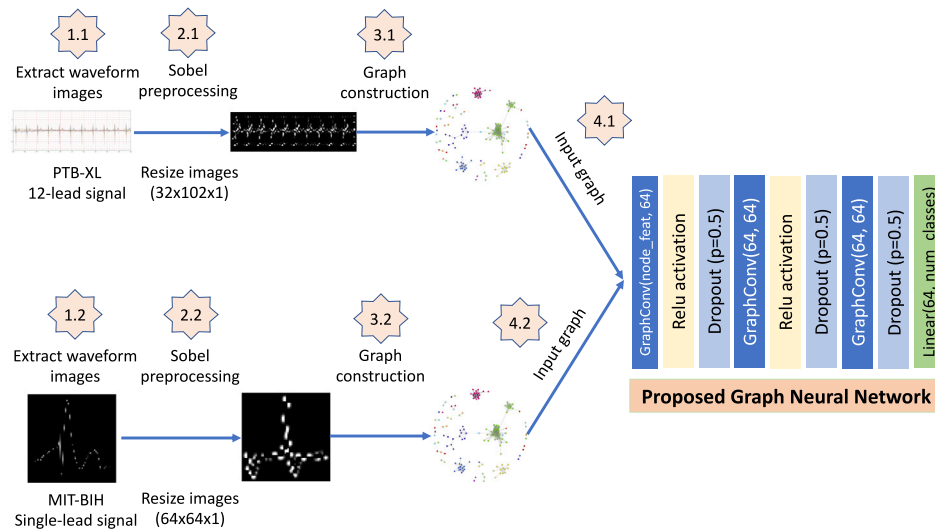


Fig. 1. Architecture of the proposed ECG-GNN model.

Table 2
Number of parameters, size, and training time.

No.	Backbone	# Params	Model size	Training time (h)	
				MIT-BIH dataset	PTB-XL dataset
1	GAT	1768	10,311	Less than 0.5	Less than 1.0
2	GATv2	3144	16,736	Less than 0.5	Less than 1.0
3	GCN	1064	1640	Less than 0.5	Less than 1.0
4	GraphConv	2952	14,479	Less than 0.5	Less than 1.0

Table 3
Parameters of the proposed ECG-GNN approach.

Stage	Name	Value/Description
1	GAT/GATv2/GCN/GraphConv	Numbers of node features
2	ReLU activation	Not applicable
3	Dropout	0.5
4	GAT/GATv2/GCN/GraphConv	32 -> 32
5	ReLU activation	Not applicable
6	Linear	32 -> number classes

hours on the two experimental datasets. Among others, GATv2 contains the largest number of parameters compared to the other models. In particular, its number of parameters is 3144, while GraphConv has 2952, and both GAT and GCN have a small number of parameters, i.e., 1768 and 1064, respectively. While these backbones are different in model size and the number of parameters, their training time is both less than 0.5 s for the MIT-BIH dataset and 1 s for the PTB-XL dataset.

Table 3 summarizes information of our proposed Graph Neural Network model. Input and output channels at the first GNN layer are equal to the number of node features of relevant input data and 64. The next two input and output channels of the other GNN layers are set as 64. Following two first GNN layers, there are activation function of ReLU (Agarap, 2018) and Dropout (Srivastava et al., 2014) of 0.5, whereas the last GNN layer is connected to ReLU activation and a linear layer with input of 64 and output as numbers of categories.

To foster follow-up research, we published the code and data together with the model's trained weights in the supporting repository.³

5. Evaluation

In this section, we present the research objectives, the datasets used for conducting an empirical evaluation on the proposed approach to

study its performance as well as to compare it with state-of-the-art baselines.

5.1. Research questions

Through a series of experiments, we study our proposed approach by answering the following research questions:

- **RQ₁**: How effective and efficient is the proposed model when performing on the 1-lead MIT-BIH and standard 12-lead superclass PTB-XL databases? We executed edge detection to transform the input images into graph-structured data using the Sobel algorithm. Afterwards, we ran the GNN model on the MIT-BIH and PTB-XL databases to study the accuracy as well as to measure how fast the model can produce predictions.
- **RQ₂**: How does our proposed approach compare to the considered baselines? We are interested in understanding if our proposed approach outperforms a state-of-the-art baseline built on top of a conventional Convolutional Neural Network combined with Long-Short Term Memory Neural Network for ElectroCardioGram classification.

5.2. Datasets

In this study, two databases namely MIT-BIH Arrhythmia⁴ (Goldberger et al., 2000) and PTB-XL (Wagner et al., 2020) are used for the evaluation. The MIT-BIH Arrhythmia database has been recorded from 47 subjects, consisting of analog 48-hour extraction of two-lead ambulatory ECG signals, and then being converted to digital form with 360 samples per second per lead with 11bit resolution over a 10 mV range. Among others, 23 recordings were randomly taken from inpatients and outpatients; the remaining 25 signals were collected from the same set but clinically significant arrhythmias were detected and annotated by cardiologists at Boston's Beth Israel Hospital between 1975 and 1980. Afterward, the signal was transformed to waveform following their annotation, and split into nine categories, i.e., Atrial Premature Contraction (APC), Left Bundle Branch Block (LBB), Right Bundle Branch Block (RBB), Paced Beat (PAB), Premature Ventricular Contraction (PVC), Ventricular Escape Beat (VEB), Ventricular Flutter Wave (VFW) and Normal (NORM). Finally, we obtained 107,620 signal images, which were then split into training/validation and testing sets as shown in Table 4.

³ <https://github.com/linhduongtuan/GraphECGNet>.

⁴ <https://www.physionet.org/content/mitdb/1.0.0/>.

Table 4
Summary of the MIT-BIH Arrhythmia database.

Category	Training set	Validation set	Testing set	Total
APC	1018	254	1272	2544
LBB	3229	807	4036	8072
NORM	30,006	7502	37,508	75,016
PAB	2810	702	3512	7024
PVC	2852	713	3565	7130
RBB	2902	726	3628	7256
VEB	42	11	53	106
VFW	189	47	236	472
Total	43,048	10,762	53,810	107,620

Table 5
Summary of the PTB-XL database.

Category	Training set	Validation set	Testing set	Total
CD	9846	2461	8206	20,513
HYP	101,658	25,414	84,716	211,788
MI	21,990	5497	18,327	45,814
NORM	551,870	137,968	333,446	1,023,284
STTC	4176	1044	3482	8702
Total	689,540	172,384	448,177	1,310,101

Table 6
Summary of the preprocessed data.

Dataset	MIT-BIH dataset	PTB-XL dataset
# of graphs	105,743	1,310,101
# of features	9	6
# of classes	8	5
# of nodes	173	135
# of edges	708	144
Degree	4.09	1.07
Dataset size (MB)	41.1	92,500
Preprocessed data size (MB)	123.6	5370
Graph-structure data size (MB)	4000	15,340
Preprocessing time (h)	2	7

The 12-lead PTBXL database was collected by the Physikalisch Technische Bundesanstalt (PTB) from 1989 to 1996. It consists of 18,885 subjects with 21,837 clinical 12-channel ECG recordings of 10 s. Since it is a huge dataset, we partly extracted it to 12-lead waveform images following settings of “validated_by_human” of “True” and sampling frequency of 100 Hz. We compose a waveform imaging dataset of 5 superclasses with Conduction Disturbance (CD), Hypertrophy (HYP), Myocardial Infarction (MI), Normal (NORM), and STT-Change (STTC). Eventually, we obtained 1,310,101 images of 10-second waveform belonging to the five categories.

The distribution of each category and training/validation/testing sets is shown in Table 5.

Table 6 depicts information about graph-structure data transformed from the databases. Since the PTB-XL dataset is much larger compared to MIT-BIH, so is its number of graphs, features, nodes, and edges. For instance, for the PTB-XL dataset, there are a total of 1,310,101 graphs, meanwhile the corresponding number for the MIT-BIH dataset is 105,743.

Examples of images from the MIT-BIH and PTB-XL datasets and their corresponding edges extracted using the Sobel filter are shown in Figs. 2 and 3, respectively. For the MIT-BIH dataset, abnormal signals were extracted using single-head based on the metadata attached to the dataset. This is done following the annotation given in the metadata. As a result, the resulting images are of square matrices. In contrast, the images in the PTB-XL dataset are in the rectangular form since they were collected in a different way, i.e., the signals were sampled using a window with length larger than 10 s. Correspondingly, in this work, we applied a window length of 10 s for getting the data.

5.3. Settings

All the experiments were performed with PyTorch version 1.10,⁵ PyG version 2.0.4.⁶ The preprocessing phases have been executed with a Mac Air M1 2020 with 16G RAM. Training and inference experiments were conducted using a workstation with configurations of operation systems Ubuntu 20.04 LTS, 96G RAM, and 24G VRAM of NVIDIA GeForce RTX 3090 with CUDA version 11.3. Our proposed GNN model was trained with 100 epochs, batch sizes of 4048, learning rate of 0.001, optimization of ADAM (Kingma & Ba, 2017).

5.4. Evaluation metrics

Given a set of images, there are the corresponding labels, i.e., $G = (G_1, G_2, \dots, G_N)$. By running a classifier, we obtain a set of predicted labels, i.e., $C = (C_1, C_2, \dots, C_N)$. The prediction performance is measured by comparing the two sets using the following metrics: *accuracy*, *precision*, *recall*, and F_1 score, being specified as follows.

Accuracy: It is the ratio of number of correct predictions to the total number items.

$$accuracy = \frac{\sum_i^N match_i}{\sum_i^N |G_i|} \times 100\% \quad (6)$$

Precision and Recall: *Precision* is the number correctly predicted instances, whereas *recall* evaluates the ability to find all relevant instances in the dataset.

$$precision_i = \frac{match_i}{|C_i|} \quad (7)$$

$$recall_i = \frac{match_i}{|G_i|} \quad (8)$$

F_1 score (F-Measure): F_1 -score is computed as the harmonic mean of Precision and Recall as follows.

$$F_1 = \frac{2 \cdot precision_i \cdot recall_i}{precision_i + recall_i} \quad (9)$$

6. Experimental results

We present and analyze the experimental results by answering the research questions introduced in Section 5.1.

6.1. RQ₁: *How effective and efficient is the proposed model when performing on the 1-lead MIT-BIH and standard 12-lead superclass PTB-XL databases?*

This research question aims to study both the *effectiveness* and *timing efficiency* of the proposed GNN model on the MIT-BIH and PTB-XL datasets. We believe that it is important to have real-time diagnosis as this helps to reduce time cost in urgent cases, providing more essential information for the patient under consideration. With the real-time diagnosis provided by the machine learning framework, doctors can have a quick overview of the situation, before proceeding with the next steps. On the other hand, accuracy is important in the context of diagnosis using Deep Learning algorithms. This is what we attempt to achieve in this paper by proposing an effective framework for pre-screening phases with a high prediction accuracy. In particular, we employed the Sobel operator to extract the most important features of images, so as to reduce the computational complexity, as well as to increase the overall prediction performance.

⁵ <https://pytorch.org>.

⁶ https://github.com/pyg-team/pytorch_geometric.

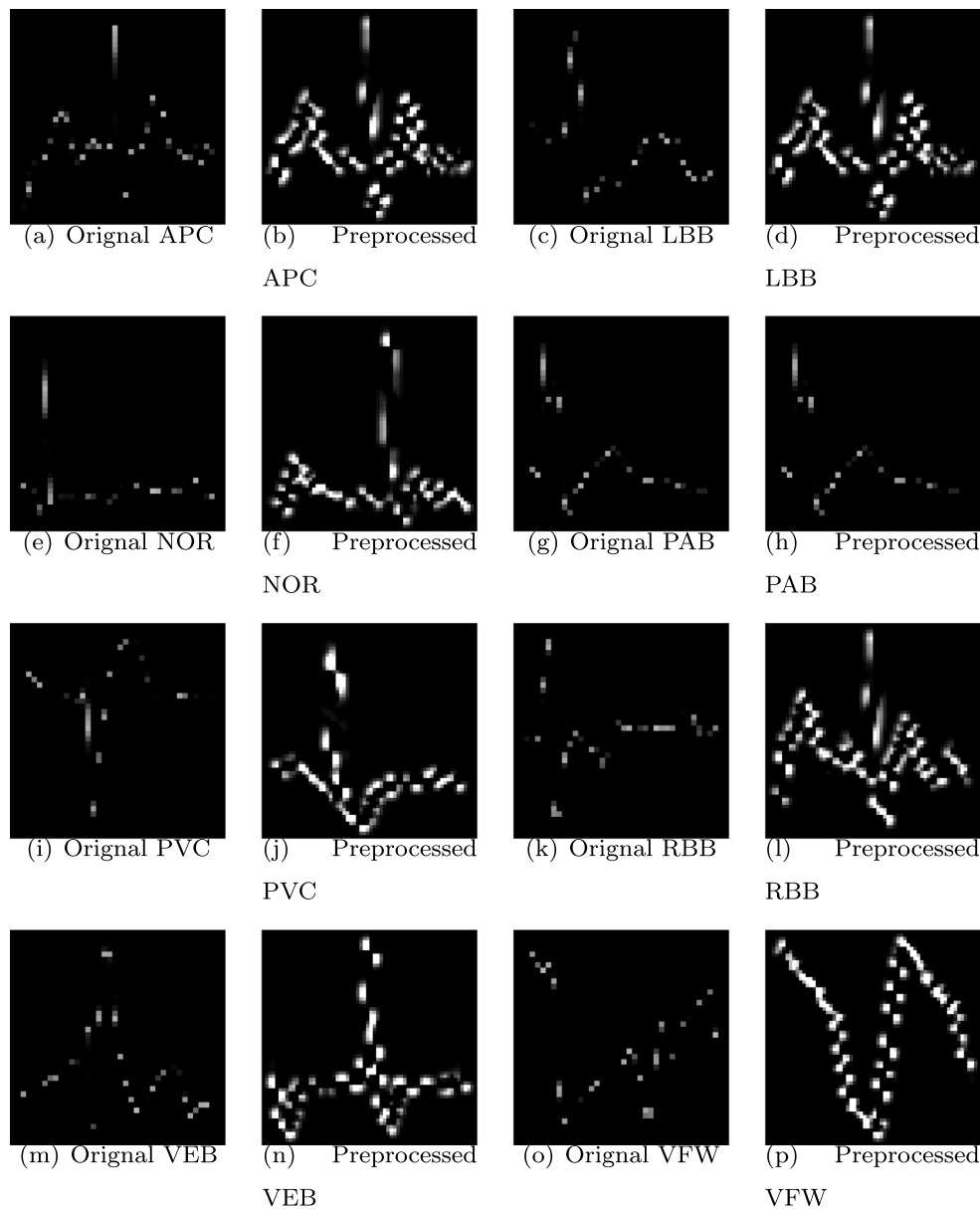


Fig. 2. Example images from the MIT-BIH dataset and their extracted edges.

6.1.1. Effectiveness

Figs. 4 and 5 display the confusion matrices by running our proposed ECG-GNN model on the given datasets. To reduce the training cost, we use an early stopping function with a setting of 10 epochs. Fig. 4 shows the achieved confusion matrices on the MIT-BIH dataset. It consists of four sub-figures which are presented for different configurations of employing layers, regarded as GAT, GATv2, GCN and GraphConv, respectively. It is clear that our proposed approach achieves an optimal accuracy for any categories in both datasets. In particular, the results shown in Figs. 4(a), 4(c), and 4(d) indicate that the ECG-GNN model reaches the maximum accuracy of 100% for all the eight categories in MIT-BIH. Only by the GATv2 configuration as shown in Fig. 4(b), the obtained accuracy is 99.90% with all 38 samples of VEB category being misclassified to RBB. It is worth noting that there is a significant disparity in number between each category. For example, in the MIT-BIH dataset, the largest category is NORM with 25,572 samples while the smallest category is VEB with only 38 samples. This means that our proposed model is able to deal with the data imbalance in both datasets.

The results obtained by running the proposed approach on the PTB-XL dataset are shown in Fig. 5. We witness a similar outcome to that achieved by the MIT-BIH dataset. It can be seen from this figure that the model gets a maximum accuracy, i.e., 100% by all the four configurations using GAT, GATv2, GCN, and GraphConv. The pattern of data imbalance is also observed in MIT-BIH dataset, in which NORM is the largest category with 333,446 testing images, and the smallest one is STTC with only 3482 testing images. Altogether, this shows that the ECG model works well both with a category with a lot of images, as well as with a category that contains few images.

Figs. 6(a)–6(d) present accuracy and loss values after each epoch during training process for the MIT-BIH dataset corresponding to different configurations including GAT, GATv2, GCN and GraphConv, respectively. From these sub-figures, we can observe that these values quickly reach the ideal rate after a few epochs. Since the MIT-BIH dataset has far fewer images than the remaining one, its accuracy and loss values converge more quickly, i.e., less than 10 epochs. Otherwise, the training process needs approximately 20 epochs to be completely optimized. Similarly, it is easy to recognize that Figs. 7(a)–7(d) share a similar pattern with that of Figs. 6(a)–6(d) in terms of the training set

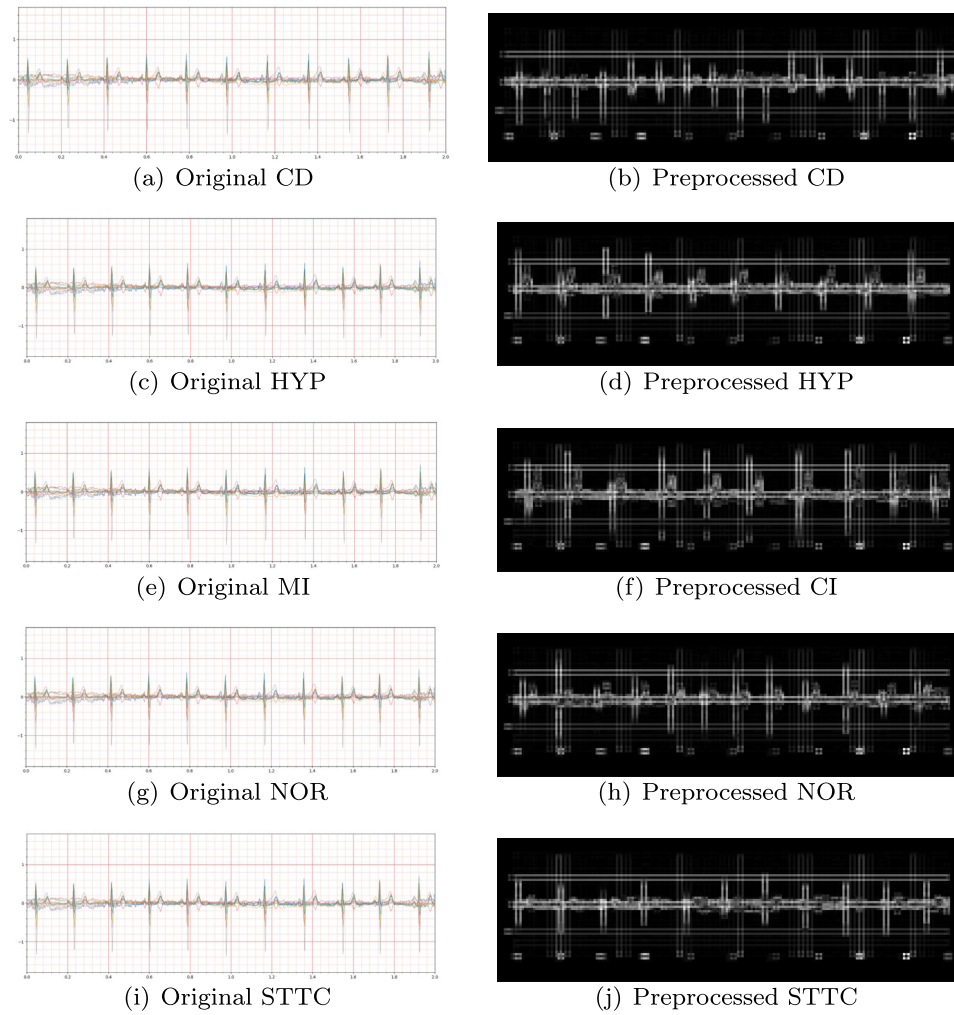


Fig. 3. Example images from the PTB-XL dataset and their extracted edges.

Table 7
Inference time.

Dataset	# test set graphs	Measured time (s)		
		Min	Average	Max
MIT-BIH	31,722	8.881	8.963	9.004
PTB-XL	448,177	49.649	49.992	50.128

and the validation set. In particular, the lines depicted for the two sets' accuracy are permanently coincided. This happens also with the lines represented for loss values. This implies that there is neither overfitting nor underfitting on the training process. Altogether, we conclude that the proposed model provides not only superior prediction results but also an ideal training performance.

6.1.2. Timing efficiency

To investigate the performance in producing diagnosis in real-time, we measured the inference time of our proposed method by running 10 times the testing phase. The resulting inference time for all the 10 rounds of execution is shown in Table 7 as the minimum, average, and maximum values in seconds for the whole testing set. Concerning the MIT-BIH dataset, where there are 31,722 testing images, our approach generates predictions very quickly, and it takes as maximum around 9 s for the whole testing set. This means that to produce prediction for an image, it needs only a very small fraction of time, becoming real-time prediction. Similarly on the PTB-XL dataset, the maximum

prediction time is 50.128 s for the whole 448,177 images. Altogether, this indicates that the proposed GNN model can provide effective prediction in real-time diagnosis.

Answer to RQ₁. Our proposed approach using GNNs is effective on the given datasets, and it yields an optimal prediction performance on the two selected datasets, converging quite quickly after few epochs. Moreover, it is also timing efficient as it can produce predictions in a real-time manner.

6.2. RQ₂: How does our proposed approach compare to the considered baselines?

There exist various approaches for the classification of electrocardiogram signals, and they mainly involve combining a conventional convolutional neural networks and long short term sequence networks in performing the training and prediction. We select three state-of-the-art models for this task from related studies, namely ResNet26D, ConvNeXt_Atto, and ConvNeXt_Femto and use them as the baseline models in the comparison with our proposed approach.

We executed the three baselines on the two given datasets, and the experimental results are collected and visualized as confusion matrices in Figs. 8, 9, and 10. Their prediction performance is analyzed as follows.

The outcomes for running ResNet26D are shown in Fig. 8, where the confusion matrices on the MIT-BIH and PTB-XL datasets are in Figs. 8(a) and 8(b), respectively. It is clear that the obtained prediction

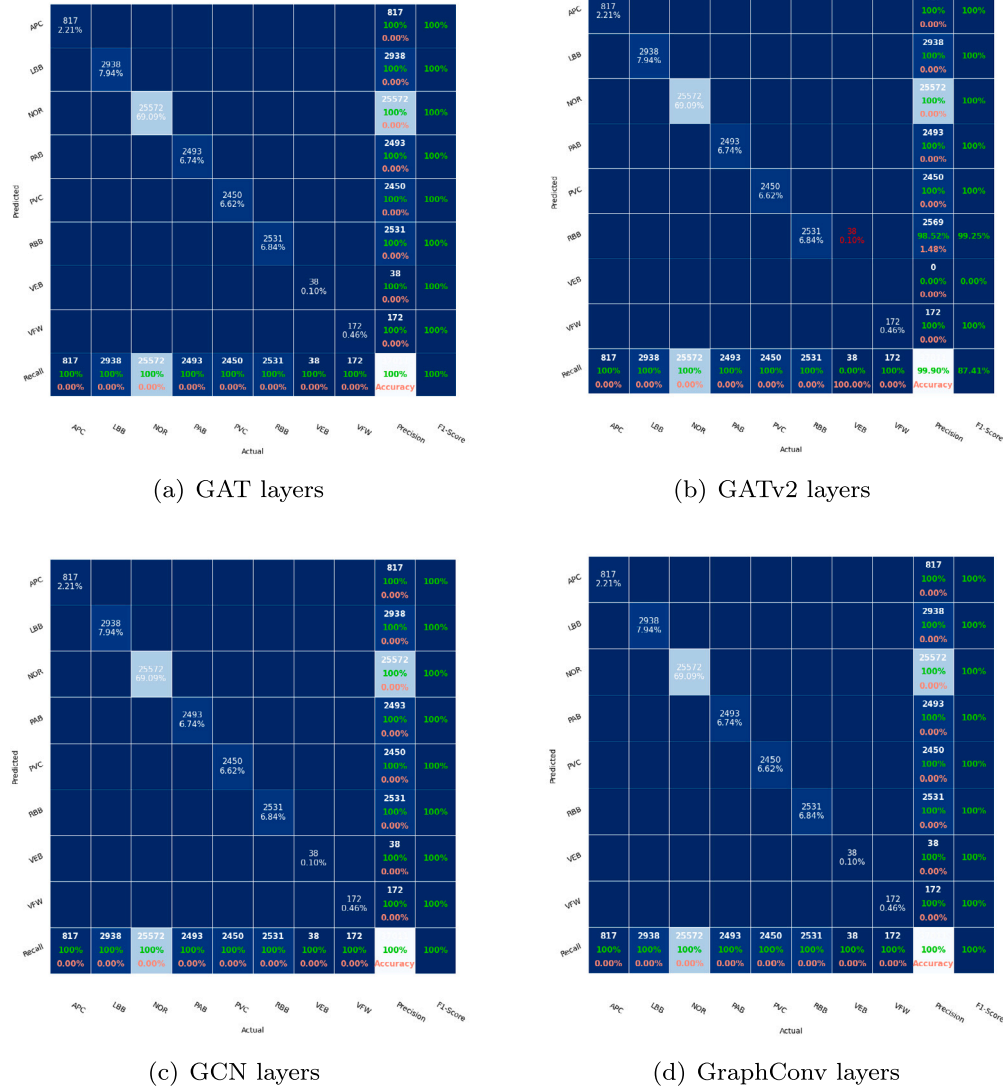


Fig. 4. Confusion matrices on the independent MIT-BIH test set.

performance is considerably low compared to that of our proposed model in Figs. 4 and 5. In particular, ResNet26D achieves a prediction accuracy of 69.72% on MIT-BIH, and it classifies almost all the samples into the NOR category, except for only 3 NOR samples labeled as PAB. For the PTB-XL dataset, the achieved accuracy is 78.11% with only 120,329 samples in the HYP category being classified correctly. The rest of this dataset is misclassified into the HYP category, implying that ResNet26D yields a low classification performance on both datasets. This is far from optimal compared to our approach using GNNs, whose the prediction accuracy is 100% by three among four configurations.

Using ConvNeXt_Atto as the classification engine, we obtain a similar outcome to that of ResNet26D. As shown in Figs. 9(a), and 9(b), the baseline gets 69.67% and 78.11% as accuracy for MIT-BIH and PTB-XL, respectively. Compared to our proposed model, ConvNeXt_Atto fails to obtain an optimal prediction performance. The results obtained for ConvNeXt_Femto on the MIT-BIH and PTB-XL datasets are shown in Figs. 10. We witness the same results as by using ConvNeXt_Atto, i.e., the accuracy is low and the baseline underperforms the GNN models.

As shown in Figs. 8(a), 9(a), and 10(a), on the MIT-BIH dataset, three baseline models classify almost all the images to the NOR category. In particular, the first baseline (ResNet26D) correctly classifies only 7500 over 10,758 images, corresponding to an accuracy of 69.72%. Apart from that, all the remaining images belong to other

categories, rather than NOR, e.g., 695 images from the PAB category are incorrectly classified into NOR. Actually, all the images in the PAB category are sorted into the wrong categories, yielding an accuracy of 0.00%. Considering the two remaining baseline models regarded as ConvNeXt_Atto and ConvNeXt_Femto, the obtained results are even worse as all images are classified into the NOR category.

Referring to the confusion matrices of the PTB dataset presented in Figs. 8(b), 9(b), and 10(b), it is clear that the three baseline models achieve similar outcomes to those on MIT-BIH. This dataset has a significant disparity between categories, in which the largest NORM category has 262,026 samples, accounting for 78.11%. The baselines return an uncertain prediction, where every image in this dataset is classified into the NORM category. That leads to the accuracy of 0% for the rest categories. The overall accuracy is significantly lower than the score of 100% obtained by our proposed model. Therefore, with the superlative accuracy for all categories, our conceived approach built on top of graph-structure data and GNN model provides significant improvement on the prediction results.

Compared with ECGNN (Zhao et al., 2022), our proposed approach is built on top of graph neural networks, employing the Sobel operator (Kanopoulos et al., 1988) and multiple GNN layers, and achieving superior performance for the same task. In particular, the maximum accuracy obtained by ECGNN (Zhao et al., 2022) is 0.94, while in

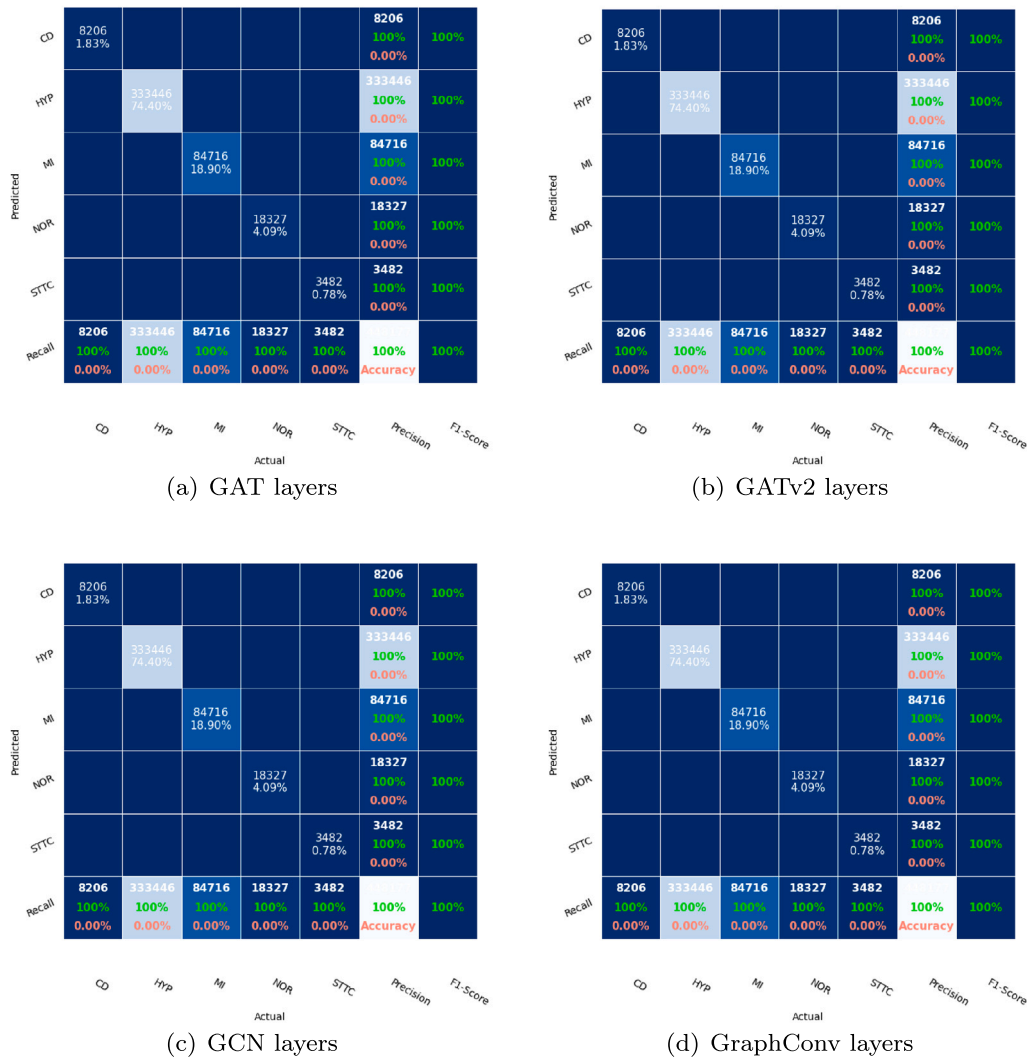


Fig. 5. Confusion matrices on the independent PTB-XL test set.

the experiments, we showed that our proposed GNN model can obtain a maximum accuracy of 100% by different configurations. Thus, we can conclude that our proposed model yields a superior performance compared to ECGNN (Zhao et al., 2022).

Answer to RQ₂. The proposed approach using GNN as the classification engine outperforms all the considered baseline models with respect to accuracy, precision, and recall.

6.3. Threats to validity

We list the threats that may influence the internal, external, and conclusion validity of our findings.

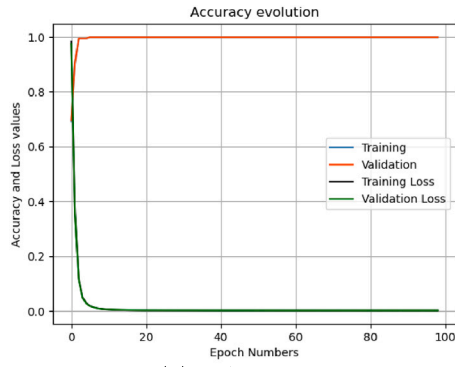
- **Internal validity.** They are the internal factors that might influence on the outcomes. A probable threat is the evaluation against the baseline approaches. Such a threat is undermined by running experiments with the original source code, as well as testing all the related tools on the collected dataset, and evaluating them with the same set of metrics.
- **External validity.** This is related to the generalizability of our conceived solution, i.e., whether it is still applicable outside the scope of this work. We anticipate that datasets collected under different conditions might have an impact on the prediction performance. Thus, training the system with data coming from different sources will enhance its generalizability in practical use.

- **Conclusion validity.** This is related to the experimental settings used in our work, concerning the simulated configurations to evaluate the approach. We performed the evaluation by means of a training set and a test set, which may not represent a real-world usage. To target a fair comparison, we made use of the same configurations to compare the systems.

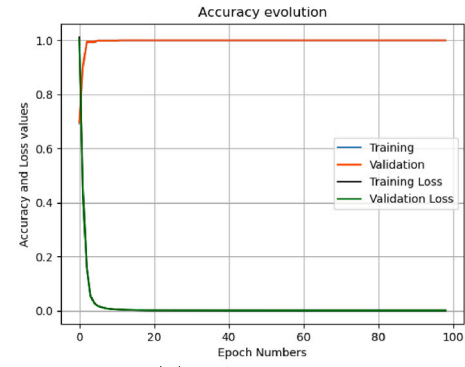
7. Conclusions and future work

This work presents a workable solution to the detection of certain CVDs from both single-lead and 12-lead ECG recordings databases. Starting from the observation that existing approaches achieved a limiting prediction performance, and most of them used Support Vector Machine algorithms, or a combination of Deep Learning models like CNN and LSTM, we hypothesize that a GNN should bring a better performance than previous existing architectures. Our proposed model has been implemented based on three layers of GraphConv and one last layer of Linear as the prediction engines. One of the main advantages of our proposed architecture is that the model outperform prediction performance compared to relevant studies in both publicly large single-lead and standard 12-lead ECG databases.

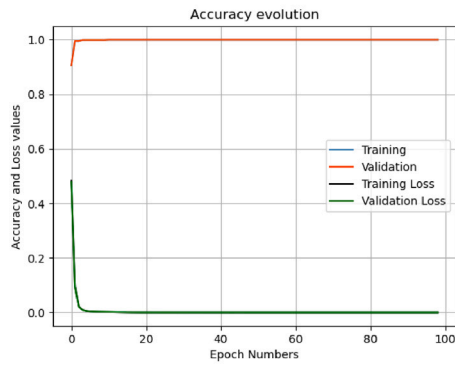
For our future work, we plan to extend the evaluation by considering other sub-classes, rhythm and form classification tasks of the PTB-XL database. This aims to better investigate intrinsic characters of the GNN model and give prominence to its novelty compared to



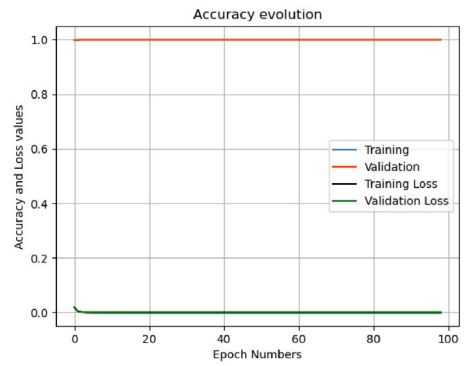
(a) GAT layers



(b) GATv2 layers

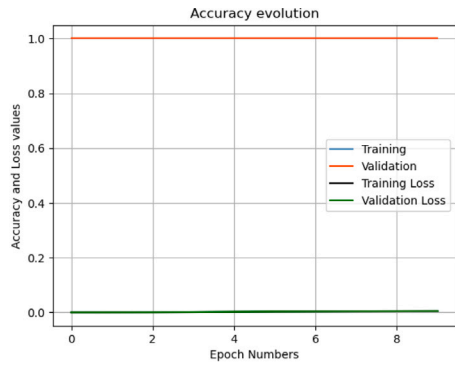


(c) GCN layers

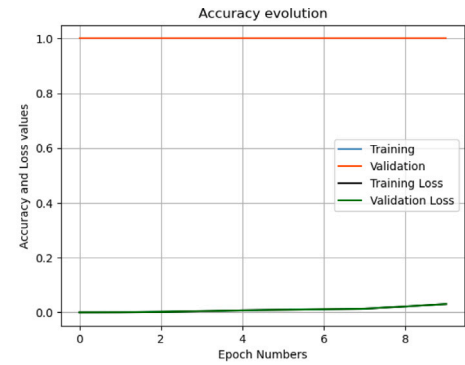


(d) GraphConv layers

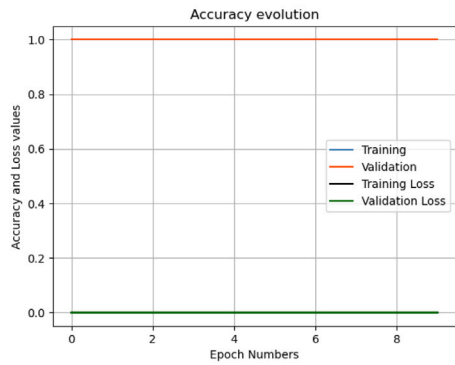
Fig. 6. Accuracy and loss on the MIT-BIH dataset.



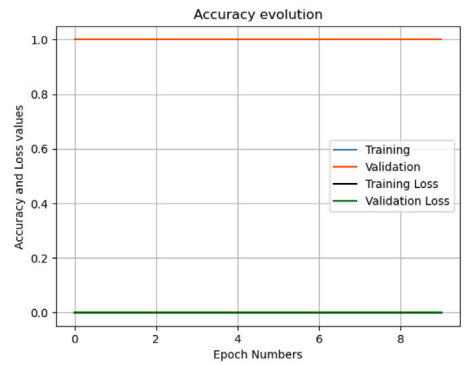
(a) GAT layers



(b) GATv2 layers

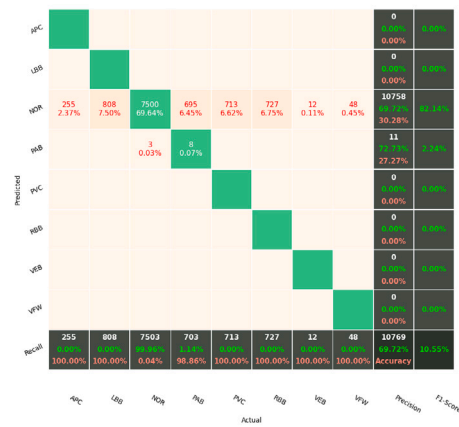


(c) GCN layers

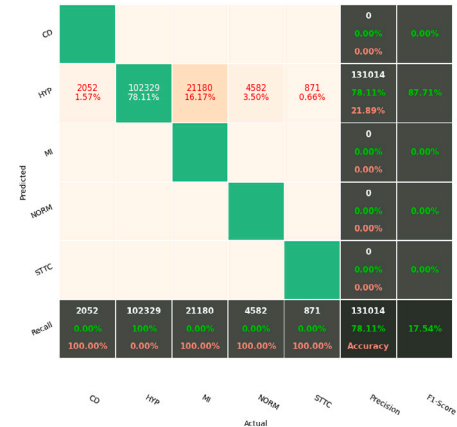


(d) GraphConv layers

Fig. 7. Accuracy and loss on the PTB-XL dataset.

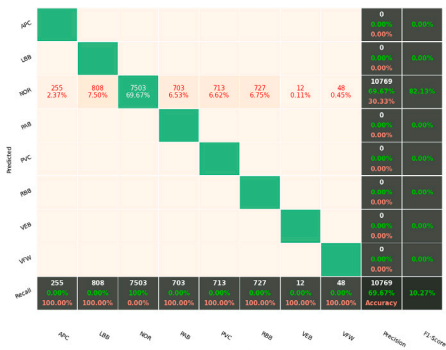


(a) MIT-BIH



(b) PTB-XL

Fig. 8. Results for ResNet26 on MIT-BIH and PTB-XL.

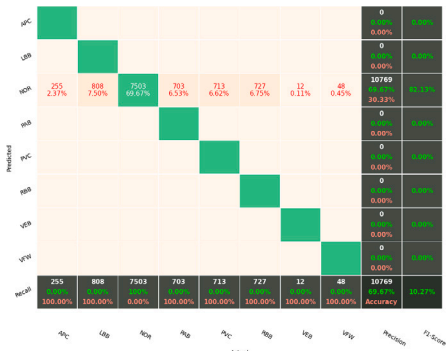


(a) MIT-BIH



(b) PTB-XL

Fig. 9. Results for ConvNeXt_Atto on MIT-BIH and PTB-XL.



(a) MIT-BIH



(b) PTB-XL

Fig. 10. Results for ConvNeXt_Femto on MIT-BIH and PTB-XL.

other ones. Last but not least, we also address to provide doctors with a practical tool to assist automatic diagnosis of ECG signals. The CAD is anticipated to work as a continuously realtime tool, being able to serve doctors at large. Such a tool is highly suitable for many wearable devices, as it helps to detect every abnormality from human ECG signals earlier and faster.

CRedit authorship contribution statement

Linh T. Duong: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Thu T.H. Doan:** Writing

– original draft, Writing – review & editing, Visualization. **Cong Q. Chu:** Writing – original draft, Data curation. **Puong T. Nguyen:** Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We would like to thank the anonymous reviewers for their comments and suggestions that help us improve our paper.

References

- Abdalla, F. Y., Wu, L., Ullah, H., Ren, G., Noor, A., & Zhao, Y. (2019). Ecg arrhythmia classification using artificial intelligence and nonlinear and nonstationary decomposition. *Signal, Image and Video Processing*, 13(7), 1283–1291.
- Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.
- Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., et al. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 25(1), 70–74.
- Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017). Heart disease and stroke statistics—2017 update: a report from the american heart association. *Circulation*, 135(10), e146–e603.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Inc: Oxford University Press, ISBN: 0198538642.
- Darmawahyuni, A., Nurmaini, S., Rachmatullah, M. N., Tutuko, B., Sapitri, A. I., Firdaus, et al. (2022). Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification. *PeerJ Computer Science*, 8, Article e825. <http://dx.doi.org/10.7717/peerj-cs.825>.
- Duong, L. T., Chu, C. Q., Nguyen, P. T., Nguyen, S. T., & Tran, B. Q. (2023). Edge detection and graph neural networks to classify mammograms: A case study with a dataset from vietnamese patients. *Applied Soft Computing*, [ISSN: 1568-4946] 134, Article 109974. <http://dx.doi.org/10.1016/j.asoc.2022.109974>, URL <https://www.sciencedirect.com/science/article/pii/S1568494622010237>.
- Duong, L. T., Le, N. H., Tran, T. B., Ngo, V. M., & Nguyen, P. T. (2021). Detection of tuberculosis from chest x-ray images: Boosting the performance with vision transformer and transfer learning. *Expert Systems with Applications*, 184, Article 115519.
- Duong, L. T., Nguyen, P. T., Iovino, L., & Flammini, M. (2023). Automatic detection of covid-19 from chest x-ray and lung computed tomography images using deep neural networks and transfer learning. *Applied Soft Computing*, [ISSN: 1568-4946] 132, Article 109851. <http://dx.doi.org/10.1016/j.asoc.2022.109851>, URL <https://www.sciencedirect.com/science/article/pii/S1568494622009000>.
- Fu, Z., Hong, S., Zhang, R., & Du, S. (2021). Artificial-intelligence-enhanced mobile system for cardiovascular health management. *Sensors*, 21(3), 773.
- Garcia, V., & Bruna, J. (2017). Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043.
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23), 215–220.
- Hannun, A. Y., Rajpurkar, P., Haghighpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., et al. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65–69.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition.
- Kanopoulos, N., Vasanthavada, N., & Baker, R. (1988). Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2), 358–367.
- Kashou, A., Ko, W.-Y., Attia, Z., Cohen, M., Friedman, P., & Noseworthy, P. (2020). A comprehensive artificial intelligence-enabled electrocardiogram interpretation program. *Cardiovascular Digital Health Journal*, 1(2), 62–70.
- Kingma, D., & Ba, J. (2017). Adam: A method for stochastic optimization.
- Le, T.-T., Lin, C.-Y., & Piedad, E. (2019). Deep learning for noninvasive classification of clustered horticultural crops – a case for banana fruit tiers. *Postharvest Biology and Technology*, [ISSN: 0925-5214] 156, Article 110922. <http://dx.doi.org/10.1016/j.postharvbio.2019.05.023>, URL <https://www.sciencedirect.com/science/article/pii/S0925521419301048>.
- Leman, A., & Weisfeiler, B. (1968). A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tekhnicheskaya Informatsiya*, 2(9), 12–16.
- Li, Q., Han, Z., & Wu, X.-M. (2018). Deeper insights into graph convolutional networks for semi-supervised learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., et al. (2018). An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics*, 8, 1368–1373. <http://dx.doi.org/10.1166/jmihi.2018.2442>.
- Loukas, A. (2020). What graph neural networks cannot learn: depth vs width. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April (2020)* 26–30. OpenReview.net, URL <https://openreview.net/forum?id=B112bp4YwS>.
- Marino, K., Salakhutdinov, R., & Gupta, A. (2017). The more you know: Using knowledge graphs for image classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 20–28). <http://dx.doi.org/10.1109/CVPR.2017.10>.
- Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., et al. (2018). Weisfeiler and leman go neural: Higher-order graph neural networks. CoRR, abs/1810.02244, URL <http://arxiv.org/abs/1810.02244>.
- Noseworthy, P. A., Attia, Z. I., Brewer, L. C., Hayes, S. N., Yao, X., Kapa, S., et al. (2020). Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ecg analysis. *Circulation: Arrhythmia and Electrophysiology*, 13(3), Article e007988.
- NT, H., & Maehara, T. (2019). Revisiting graph neural networks: All we have is low-pass filters.
- Oono, K., & Suzuki, T. (2021). Graph neural networks exponentially lose expressive power for node classification.
- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, [ISSN: 0957-4174] 97, 205–227. <http://dx.doi.org/10.1016/j.eswa.2017.12.020>, URL <http://www.sciencedirect.com/science/article/pii/S0957417417308333>.
- Rasti-Meymandi, A., & Ghaffari, A. (2022). A deep learning-based framework for ecg signal denoising based on stacked cardiac cycle tensor. *Biomedical Signal Processing and Control*, [ISSN: 1746-8094] 71, Article 103275. <http://dx.doi.org/10.1016/j.bspc.2021.103275>, URL <https://www.sciencedirect.com/science/article/pii/S1746809421008727>.
- Saha, P., Mukherjee, D., Singh, P. K., Ahmadian, A., Ferrara, M., & Sarkar, R. (2021). Graphcovidnet: A graph neural network based model for detecting covid-19 from ct scans and x-rays of chest. *Scientific Reports*, 11(1), 1–16.
- Sellami, A., & Hwang, H. (2019). A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Systems with Applications*, [ISSN: 0957-4174] 122, 75–84. <http://dx.doi.org/10.1016/j.eswa.2018.12.037>, URL <http://www.sciencedirect.com/science/article/pii/S0957417418308054>.
- Smith, J. M., & Jones, A. B. (2012). *Book title* (7th ed.). Publisher.
- Soltanisehat, L., Alizadeh, R., Hao, H., & Choo, K.-K. R. (2020). Technical, temporal, and spatial research challenges and opportunities in blockchain-based healthcare: A systematic literature review. *IEEE Transactions on Engineering Management*, 1–16. <http://dx.doi.org/10.1109/TEM.2020.3013507>.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Svozil, D., Kvasnicka, V., & Pospíchal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39, 43–62. [http://dx.doi.org/10.1016/S0169-7439\(97\)00061-0](http://dx.doi.org/10.1016/S0169-7439(97)00061-0).
- Tison, G. H., Sanchez, J. M., Ballinger, B., Singh, A., Olgin, J. E., Pletcher, M. J., et al. (2018). Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiology*, 3(5), 409–416.
- Wagner, P., Strödtz, N., Bousselet, R.-D., Kreisler, D., Lunze, F. I., Samek, W., et al. (2020). Ptb-xl, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1), 1–15.
- Wang, Z., Chen, T., Ren, J., Yu, W., Cheng, H., & Lin, L. (2018). Deep reasoning with knowledge graph for social relationship understanding. arXiv preprint arXiv:1807.00504.
- Wang, X., Ye, Y., & Gupta, A. (2018). Zero-shot recognition via semantic embeddings and knowledge graphs. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 6857–6866). <http://dx.doi.org/10.1109/CVPR.2018.00717>.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. <http://dx.doi.org/10.1109/TNNLS.2020.2978386>.
- Zhang, G., Eddy Patuwo, B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, [ISSN: 0169-2070] 14(1), 35–62. [http://dx.doi.org/10.1016/S0169-2070\(97\)00044-7](http://dx.doi.org/10.1016/S0169-2070(97)00044-7), URL <https://www.sciencedirect.com/science/article/pii/S0169207097000447>.
- 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). <http://dx.doi.org/10.1109/BIBM55620.2022.9995419>.
- Zhu, H., Cheng, C., Yin, H., Li, X., Zuo, P., Ding, J., et al. (2020). Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. *The Lancet Digital Health*, 2(7), e348–e357.