A Weighted Graph Attention Network Based Method for Multi-label Classification of Electrocardiogram Abnormalities

Hongmei Wang, Wei Zhao, Zhenqi Li, Dongya Jia, Cong Yan, Jing Hu, Jiansheng Fang, and Ming Yang

Abstract— The multi-label electrocardiogram (ECG) classification is to automatically predict a set of concurrent cardiac abnormalities in an ECG record, which is significant for clinical diagnosis. Modeling the cardiac abnormality dependencies is the key to improving classification performance. To capture the dependencies, we proposed a multi-label classification method based on the weighted graph attention networks. In the study, a graph taking each class as a node was mapped and the class dependencies were represented by the weights of graph edges. A novel weights generation method was proposed by combining the self-attentional weights and the prior learned co-occurrence knowledge of classes. The algorithm was evaluated on the dataset of the Hefei Hi-tech Cup ECG Intelligent Competition for 34 kinds of ECG abnormalities classification. And the *micro-f*1 and the *macro-f*1 of cross validation respectively were 91.45% and 44.48%. The experiment results show that the proposed method can model class dependencies and improve classification performance.

I. INTRODUCTION

Electrocardiogram (ECG) is widely used for prescreening and physical examination of heart disease. Cardiac abnormalities often occur simultaneously, so the same ECG record has multiple labels at the same time. Discriminating the abnormal type shown in the ECG signal can assist the diagnosis of heart disease. However, in the hospital, the huge amount of ECG makes clinicians' time-consuming identification. Besides, due to the coexistence of multiple abnormalities, multi-label ECG signals are complex and difficult to identify. Therefore, the automatic analysis algorithm of multi-label ECG is of great significance for the detection of heart disease.

In the last decades, a variety of methods have been proposed for the classification of ECG signals. Methods combining ECG features and machine learning algorithm[1] as well as deep learning methods[2] are applied to classify the record into one of the ECG abnormalities, such as atrial fibrillation, myocardial infarction, et al. These methods mainly focus on multi-class ECG classification. However, in reality, an ECG record may contain multiple concurrent abnormalities, which is a multi-label classification problem. It is more complicated and more difficult to distinguish the concurrent abnormalities. Recently, there are also some methods for the multi-label analysis of ECG. Luo et al.[3] used the binary relevance method for multi-label classification, which is limited to ignore the correlations between classes and looks at each class independently. Our previous work[4] used the sequence generation module (SGM), which predicted labels in a sequential fashion, based on frequency or pre-defined orders. Besides, the number of classes detected by these methods is much less than the number of actual cardiac abnormalities.

In this paper, we proposed an algorithm based on a multi-label weighted graph attention network for ECG classification (MLWGAT). The method utilized the graph attention network to model the dependencies of cardiac abnormalities. A novel weight scheme was proposed to improve the original graph attention network (GAT). Our method was evaluated on the dataset provided by the Hefei Hi-tech Cup ECG Intelligence Competition (HFECGIC) for the multi-label classification of 34 kinds of abnormalities. The results showed that the proposed MLWGAT can be used for modeling class correlations thus improved classification performance.

II. METHODS

The architecture of the proposed MLWGAT is shown in Fig. 1. The method consists of three modules. In the first module, features of ECG are extracted from a convolutional neural network (CNN). A graph network is adopted in the second module to model the class dependencies. In the last module, features generated by the previous modules are combined to predict the probability of each class.

A. ECG Representation Learning

The goal of the ECG representation is to learn the features of each class (abnormality). As shown in Fig. 1, a resnet scheme[5] is adopted. Different from the original resnet, all 2-dimensional operations are replaced by 1-dimensional ones since ECG is the 1-dimensional signals. Besides, a convolutional layer is adopted otherwise a fully-connected (FC) layer as the output layer to obtain the representation of each class.

B. Weighted Graph Attention Network (WGAT)

Graph attention networks [6] deals with graph-structured data. In the paper, we improve the GAT by integrating a co-occurrence weight to the masked attentional weights to model the class dependencies, which is named the weighted graph attention network.

First, we construct graphs by taking class representation as nodes. As shown in Fig. 1, the features of each

Hongmei Wang, Wei Zhao, Zhenqi Li, Dongya Jia, Cong Yan, Jing Hu, Jiansheng Fang and Ming Yang are with the Central Research Institute for Guangzhou Shiyuan Electronics Co., Ltd, 510530, Guangzhou, China. wanghongmei@cvte.com

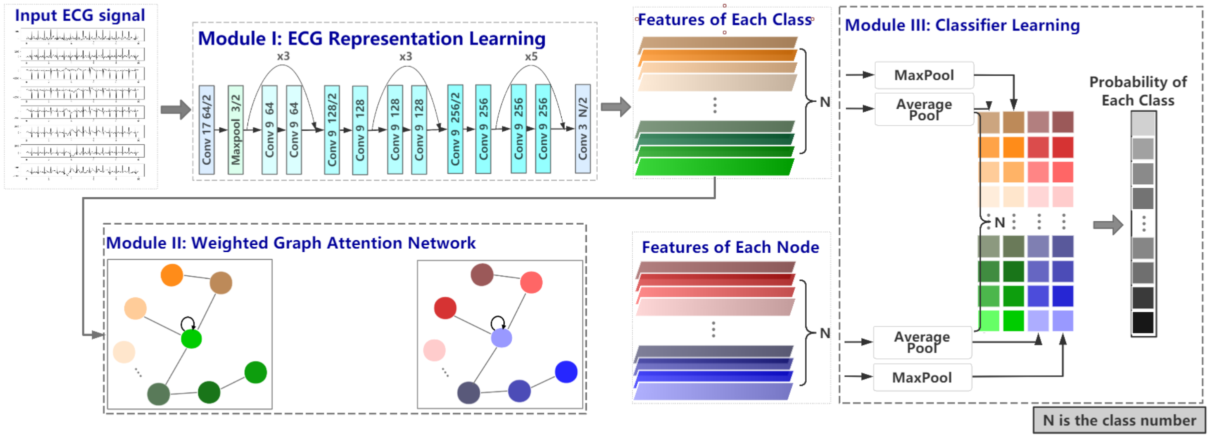

Fig. 1: Framework of the Proposed MLWGAT Model for Multi-label Classification of ECG

class output by the ECG representation module are the inputs of the WGAT. Then normalized self-attentional weights $\alpha = \{\alpha_{ij}\}_{i,j=1,1}^{N,N}$ were calculated

$$\alpha_{ij} = softmax_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k\in\mathcal{N}_i} exp(e_{ik})} \tag{1}$$

where α_{ij} means the normalized attentional coefficients of node i to node j and $\mathcal{N}_i$ means the neighborhood of node i. e_{ij} is the self-attentional weight of node i to node j. The detail of e_{ij} can be referenced in [6]. In the graph, for a class (node), only limited classes co-occur with it. So only attentions to the related neighborhoods should be considered when calculating the normalized attention coefficients.

The graph structure information is injected into the mechanism of the WGAT by the set $\mathcal{N}_i$, so how to get the graph structure, i.e. nodes and edges between nodes, is a key point. In our paper, we propose a novel weight scheme to improve the original GAT via mining the class co-occurrence patterns within the data set. We obtain the co-occurrence matrix of all classes, denoted as $\mathrm{M} = \{m_{ij}\}_{i,j=1,1}^{N,N}$, where N is the class number and m_{ij} is the number of records that have abnormalities of class i and class j simultaneously. The number of records with class i is denoted as l_i. Then, the co-occurrence probability matrix $\mathrm{P} = \{p_{ij}\}_{i,j=1,1}^{N,N}$, is calculated

$$p_{ij} = p_{ij}/l_i \tag{2}$$

p_{ij} means the probability that class j appears when class i exists. So for node i, if $p_{ij} > 0$, it means that class i and j is possible to occur on the same ECG record. So in this case, we think node j is the neighborhood of node i. In this way, it is obvious that $\mathcal{N}_i = \{j|p_{ij} > 0\}$.

For ECG abnormalities, there are certain rules about which heart diseases will occur simultaneously, which could be reflected from the co-occurrence probability matrix to a certain extent. So on basis of the self-attentional weights α, we use the co-occurrence probability matrix to weight the α. The weight of the WGAT is denoted as β, $\beta \in R^{N\times N}$

$$\beta = \alpha \cdot P \tag{3}$$

Like [6], a k-head attention scheme is also implemented in WGAT. The output of the k-head attention in WGAT is

$$\overrightarrow{h_i} = \|_{k=1}^{K} \sigma(\sum_{j\in\mathcal{N}_i} \beta_{ij}^k W^k \overrightarrow{h_j}) \tag{4}$$

where $\overrightarrow{h_j}, \overrightarrow{h_i}$ and W^k have the same meaning as in [6]. And $||$ is the concatenated operation. σ is the exponential linear units (ELU) activation function.

The weighted graph attention network helps construct the true relations of ECG abnormalities. In the second module, we apply a 4-head weighted attention network computing 64x4 features for N nodes, where N is the class number. After that, a convolutional network and a pooling layer are adopted to reduce redundancy.

C. Classifier Learning

Both global average pooling and global max-pooling are used for the features learned by the CNN and the WGAT. All four parts of features are concatenated and followed by an FC layer to output the probability of each class. If the probability of class i is larger than 0.5, it predicts that the record has abnormalities of class i.

III. EXPERIMENTS

A. Dataset and Experiment Settings

We used the dataset of the second round contest of the HFECGIC [7] for a multi-label classification task of 34 classes, as shown in the table. I. The dataset contains 20036 subjects and each subject is represented by one 10-second and eight-lead (i, ii, v1, v2, v3, v4, v5, v6) record. The sampling frequency is 500 Hz. Each ECG record was mean normalized and resampled to 204.8 Hz.

The output probability of each class by the MLWGAT is noted as $\hat{y}$. Assuming that the ground truth of the ECG record is $\mathrm{y} = \{y^i\}_{i=1}^{N}$. y^i is a value of 0 or 1. When y^i is 1, it means that the abnormality of class i exists

TABLE I: Record Number of Each Class

Class	Record number(%)	Class	Record number(%)	Class	Record number(%)
Sinus bradycardia(SB)	5264(26.3)	Atrial premature beats (APB)	314(1.57)	Non-specific T-wave abnormalities (NS-T)	34(0.17)
Sinus rhythm (SR)	9501(47.42)	ST-T change	299(1.49)	Right atrium enlargement (RAE)	32(0.16)
Sinus tachycardia (ST)	4895(24.43)	ST segment change (ST change)	286(1.43)	Rapid ventricular rate (RVR)	29(0.14)
T wave change (T change)	3479(17.36)	One degree atrioventricular block (I AVB)	142(0.71)	Complete left bundle branch block (CLBBB)	25(0.12)
Left axis deviation (LAD)	1124(5.61)	Incomplete right bundle branch block (IRBBB)	126(0.63)	Left bundle branch block (LBBB)	25(0.12)
Right axis deviation (RAD)	1124(5.61)	Atrial fibrillation (AF)	120(0.60)	Short PR interval (sPR)	23(0.11)
Sinus arrhythmia (SA)	901(4.50)	Non-specific ST segment abnormalities (NS-ST)	64(0.32)	Early repolarization (ER)	22(0.11)
Right bundle branch block (RBBB)	551(2.75)	Counterclockwise rotation (CCR)	60(0.30)	Paced rhythms	16(0.08)
Ventricular premature beat (VPB)	543(2.71)	Abnormal Q waves (AQW)	52(0.26)	nonspecific ST-segment and T-wave abnormalities (NS-STT)	16(0.08)
Complete right bundle branch block (CRBBB)	418(2.09)	Left anterior branch block (LABB)	35(0.17)	Poorly increasing anterior wall R wave (PRWP)	16(0.08)
Left ventricular high voltage (LVHV)	414(2.07)	clockwise rotation (CR)	35(0.17)	Fusion wave	7(0.03)
				QRS low voltage (LQRSV)	3(0.01)

on the record. The MLWGAT network was trained with the following multi-label classification loss

$$\mathscr{L} = \sum_{i=1}^{N} y^i log(s(\hat{y}^i)) + (1-y^i) log(1-s(\hat{y}^i)) \qquad (5)$$

where $s(\cdot)$ is the sigmoid function. Adam [8] was used as the optimizer to train the network. The initial learning rate (LR) was 0.001. And the LR decayed by a factor of 2 every 30 epochs. Weight decay was 10^{-4}. The batch size is 32 and number of epoch is 110.

B. Experiment Results

To evaluate the proposed MLWGAT, we employed the five-fold cross validation subject-wise experiments. Here the training and test folds contained records from different subjects. The *micro-*$f1$ and *macro-*$f1$[9] was reported for performance evaluation. *macro-*$f1$ assigns the same weight to each class, whereas *micro-*$f1$ assigns the same weight to each test record.

For comparison, we did the same experiment on three baselines. The first one is the SGM-based multi-label model proposed in [4]. The binary relevance (BR) algorithm is a classic method for multi-label classification[10]. So a BR algorithm based on the resnet was taken as a baseline. It has the same resnet scheme as the module I of the MLWGAT. The difference is that the second module was removed and the input of the final FC layer does not contain the features from the WGAT network. The third baseline is the GAT-based network, which has the same structure as MLWGAT except that the WGAT module was replaced by the GAT. Paired t-tests were also conducted on the *micro-*$f1$ values and *macro-*$f1$ values to examine the MLWGAT against all the baseline methods, referenced [11].

TABLE II: Comparisons and Paired T-test Results of the Proposed MLWGAT and the Baselines

	micro-$f1$ (P value)	*macro-*$f1$ (P value)
SGM-based Method [4]	90.72%(0.008)	40.96%(<0.001)
Resnet-based BR	90.98%(0.018)	41.61%(0.009)
GAT-based Method	91.09%(0.047)	43.95%(0.045)
MLWGAT	91.45%(-)	44.48%(-)

The results of the experiment and the t-test were shown in table II. The *micro-*$f1$ and *macro-*$f1$ of the cross validation of the proposed MLWGAT are respectively 91.24% and 44.48%. Comparing with the three baselines, p values of the paired t-tests are smaller than 0.05, which indicates the significant difference between the MLWGAT and the compared baseline. The proposed MLWGAT achieved the best *micro-*$f1$ and the *macro-*$f1$ on the multi-label classification of the ECG abnormalities.

We calculated the sensitivity(SE) of each class, as shown in Fig. 2. For the MLWGAT, the SEs of the SB, ST, SR, CRBBB are relatively high. The record number of different classes is imbalanced. The record number of CR, N-ST, RAE, sPR, ER, Paced rhythms, NS-STT, PRWT, LQRSV, Fusion Wave is less than 0.17%, which results in poor SE of these classes. And also it results in low *macro-*$f1$ because the *macro-*$f1$ treats each class equally and doesn't consider data imbalance.

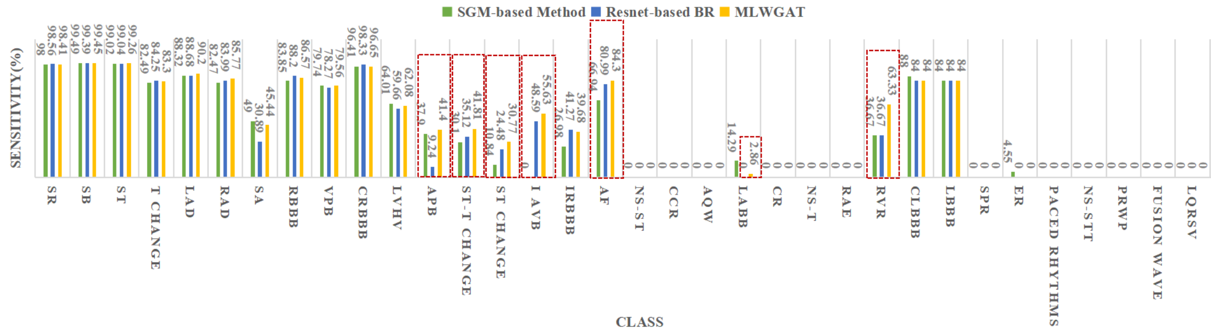

Fig. 2: Comparisons of Sensitivity Among the Proposed Method and Baselines Without Graph Structure

IV. Discussion

The graph networks naturally have advantages in characterizing node relationships. In the paper, the proposed MLWGAT takes advantage of the graph network to model the dependencies between the ECG abnormalities, which is helpful for performance improvement. So when compared to the resnet-based BR method which does not contain the WGAT module, as shown in table II, the *micro-f*1 and *macro-f*1 of the MLWGAT are higher.

Modeling the class dependencies also helps the classification of classes with few records and alleviates the impact of data imbalance. Compared with the resnet-based BR algorithm and the SGM-based method, as shown in the red boxes of the Fig. 2, the proposed MLWGAT network gets higher SE on the classes with fewer records, which is difficult to detect. And compared to the SGM model, both the proposed MLWGAT and the SGM model have considered the class dependencies. However, the MLWGAT uses graphs to directly describe the relationship between abnormalities, while SGM predicts coexisting abnormalities by indirectly learning the orders of the abnormality output the decoder.

Different from the GAT, a new weight scheme was proposed in the MLWGAT considering both the influence of the attention between nodes and the actual relationships represented by the statistics of the data set. The scheme makes good use of the prior knowledge of the class dependencies. When removing the proposed weight scheme, as shown in table II, the GAT-based method performs lower *micro-f*1 and *macro-f*1 than these of the MLWGAT, which indicates that the proposed scheme is effective.

V. Conclusion

We proposed a WGAT based network for the multi-label classification of ECG abnormalities. To mine the abnormality dependencies, a graph taking each class as a node was structured and the class dependencies were represented by the weights between nodes. To improve the classification performance, a novel weight scheme was proposed by introducing the prior class co-occurrence probability. The proposed WGAT-based end-to-end network was evaluated on a multi-label classification task for ECG abnormalities of imbalanced 34 classes. The experiment results showed that the proposed method achieved a *micro-f*1 of 91.45% and a *macro-f*1 of 44.48%. In addition, the MLWGAT outperformed the baseline methods and achieved higher SE on detection of classes with fewer samples. In the future, an interesting research direction would be introducing the few-shot or zero-shot learning method to improve the detection performance of classes with few records.

References

[1] S. Ansari, N. Farzaneh, M. Duda, K. Horan, H. B. Andersson, Z. D. Goldberger, B. K. Nallamothu, and K. Najarian, "A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records," IEEE reviews in biomedical engineering, vol. 10, pp. 264–298, 2017.

[2] P. Bizopoulos and D. Koutsouris, "Deep learning in cardiology," IEEE reviews in biomedical engineering, vol. 12, pp. 168–193, 2018.

[3] C. Luo, H. Jiang, Q. Li, and N. Rao, "Multi-label classification of abnormalities in 12-lead ecg using 1d cnn and lstm," in Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting. Springer, 2019, pp. 55–63.

[4] D. Jia, W. Zhao, Z. Li, C. Yan, H. Wang, J. Hu, and J. Fang, "An ensemble neural network for multi-label classification of electrocardiogram," in Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting. Springer, 2019, pp. 20–27.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[6] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," arXiv preprint arXiv:1710.10903, 2017.

[7] TIANCHI, "Hefei hi-tech cup ecg intelligent competition," 2019. [Online]. Available: https://tianchi.aliyun.com/competition/entrance/231754/information

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[9] T. Chen, Z. Wang, G. Li, and L. Lin, "Recurrent attentional reinforcement learning for multi-label image recognition," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[10] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," IEEE transactions on knowledge and data engineering, vol. 26, no. 8, pp. 1819–1837, 2013.

[11] G.-T. Zhou, T. Lan, W. Yang, and G. Mori, "Learning class-to-image distance with object matchings," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 795–802.