Diagnostic of Multiple Cardiac Disorders from 12-lead ECGs Using Graph Convolutional Network Based Multi-label Classification

Zheheng Jiang[1], Tiago P Almeida[1], Fernando S Schlindwein[1], G André Ng[1], Huiyu Zhou[1], Xin Li[1]

[1]University of Leicester, Leicester, UK

Abstract

Automated detection and classification of clinical electrocardiogram (ECG) play a critical role in the analysis of cardiac disorders. Deep learning is effective for automated feature extraction and has shown promising results in ECG classification. Most of these methods, however, assume that multiple cardiac disorders are mutually exclusive. In this work, we have created and trained a novel deep learning architecture for addressing the multi-label classification of 12-lead ECGs. It contains an ECG representation work for extracting features from raw ECG recordings and a Graph Convolutional Network (GCN) for modelling and capturing label dependencies. In the PhysioNet/Computing in Cardiology Challenge 2020 [1], our team, Leicester-Fox, reached a challenge validation score of 0.395, and full test score of -0.012, placing us 34 out of 41 in the official ranking.

1. Introduction

The electrocardiogram (ECG) is a clinical tool widely utilised for the clinical diagnosis of multiple cardiac diseases. The standard 12-lead ECG records resulting electrical activity of the heart collected from different angles, including six limb leads from the vertical plane and six chest leads from the horizontal plane [2]. However, manual interpretation of ECG is a time-consuming task, and requires experienced cardiologist [3]. Thus, computer-aided interpretation has become increasing in the process of clinical diagnosis, since such technique assists the cardiologist with health care decision making [3].

In traditional approaches, a variety of features are firstly extracted from ECG recordings using different techniques, such as Discrete Wavelet Transform (DWT) [4] and Pan Tompkins algorithm [5]. Then, a classification method, such as Support Vector Machine (SVM) [6], Hidden Markov model (HMM) [7] or random forests [8], is employed for classification. However, these approaches rely heavily on the carefully selected features, so is difficult to handle multi-class classification tasks using these approches [9]. Deep neural networks (DNNs) have recently achieved great success in detecting cardiovascular abnormalities from single-lead or 12-lead ECGs [10, 11]. The major advantage of DNNs is that they are able to automatically learn useful features from raw input data without requiring data preprocessing, feature engineering or handcrafted rules [10]. These methods, however, treat the problem of multiple cardiac disease recognition as a multi-class classification problem, where multiple cardiac abnormalities are regarded as mutually exclusive classes. In reality, it is possible that more than one cardiac disorders might exist concurently during the collection of ECG signals. Therefore, further work is needed to identify the correlations among labels instead of treating each label independently.

In the present work, we sought to 1) develop a novel end-to-end multi-label cardiac disease detection framework, where a deep CNN model and a bi-directional gated recurrent unit (GRU) are combined to learn high-level feature representation of ECG, and a GCN is employed to embed our label graph into inter-dependent cardiac disease classifier which is trained using our proposed class-aware Binary Cross-entropy Loss, 2) design a correlation matrix based on label dependencies to guide the information propagation among nodes in GCN, 3) demonstrate the effectiveness and efficacy of our architecture on the ECG dataset of PhysioNet/CinC Challenge 2020 [1].

2. Model Architecture

Overall framework of our deep learning model is shown in Fig.1. Our network consists of two modules for multi-label ECG diagnosis. In the first module, a deep 1D CNN followed by a bi-directional GRU layer was developed to learn ECG representations. In the second module, a three-layer GCN model was proposed to learn the inter-relationships of labels. Finally a class-aware Binary Cross-entropy Loss is proposed to jointly train both networks.

2.1. ECG representation network

In this module, we firstly applied Convolutional Neural Networks (CNNs) to learn high-level feature representa-

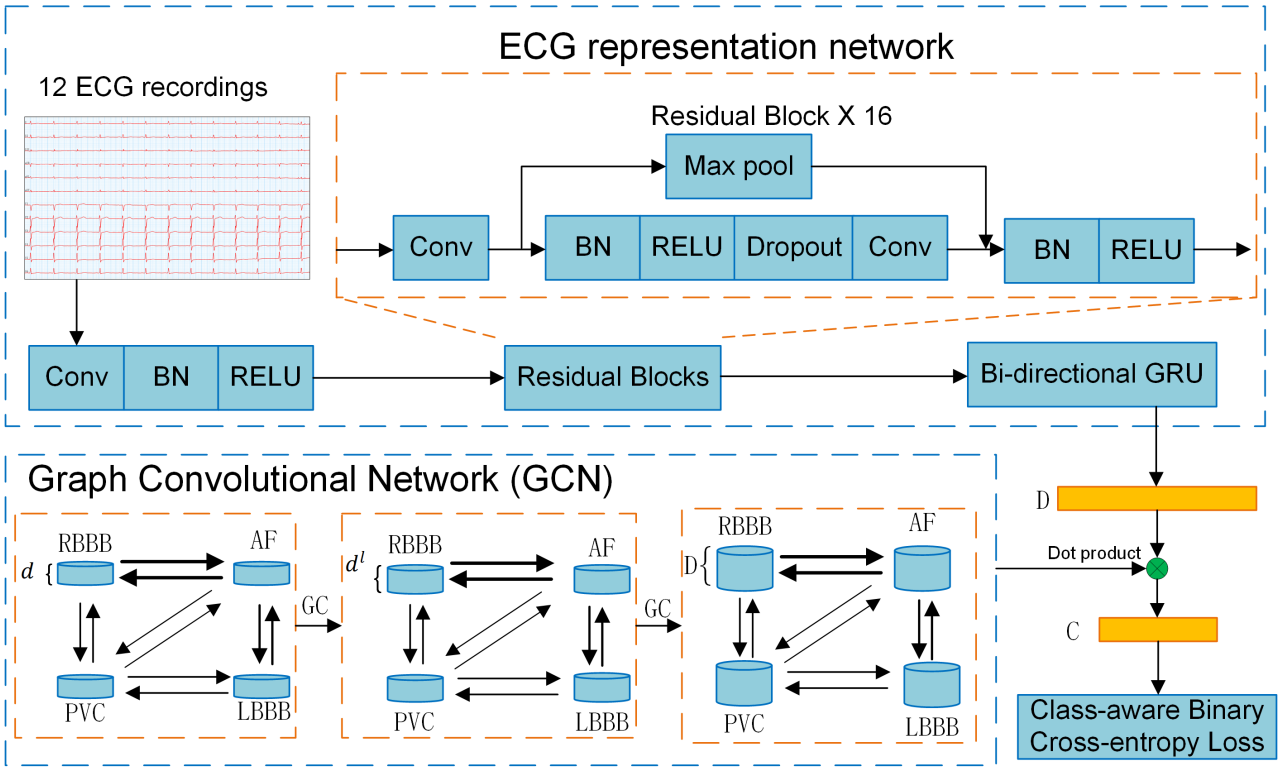

Figure 1. Overall framework of our deep learning model.

tion of ECG recordings. In order to make CNNs tractable for optimisation, a Residual Neural Network similar to [10,11] was adopted to add a shortcut connection that skips two convolutional layers. The network consists of a convolutional layer (Conv) followed by 16 residual blocks with two convolutional layers per block. The width of filters is fixed with 16 in every convolutional layer. The number of filters per convolutional layer starts with 32, and after the first four residual blocks, it doubles at the first convolutional layers in every fourth residual block. Every second residual block subsamples its inputs by a factor of 2. We applied a Batch Normalisation (BN) for rescaling the output of each convolutional layer and a rectified linear activation unit (RELU) as a nonlinear activation function. The dropout layers with a rate of 0.2 after RELU were used to prevent overfitting. A bi-directional GRU layer was finally applied to extract temporal features from the time series of CNN feature. We defined the ECG representation network as a mapping function $f_{cnn+GRU}$ and the ECG-level feature F:

$$F = f_{cnn+GRU}(E; \theta_{cnn+GRU}) \in \mathbb{R}^D \quad (1)$$

where $\theta_{cnn+GRU}$ and D denotes model parameters and the output dimension of the ECG representation network and E is an input of 12-ECG recording.

2.2. GCN based multi-label classification

A novel GCN based model has been used to capture the label correlations for multi-label classification of cardiac diseases. GCN was firstly proposed in [12] to generalise CNNs from regular domain, such as image and speech, to irregular domains, like irregular graphs. Kipf et al. [13] also introduced GCN to perform semi-supervised classification on graph-structured data, which was motivated from a first-order approximation of spectral graph convolutions. Inspired by these approaches, we sought to implement a GCN based mapping function to learn label dependencies. Unlike standard convolutions that operate on data lying on Euclidean space, the goal of GCN is to learn a function $f(\cdot,\cdot)$ of feature matrix $H^l \in \mathbb{R}^{n\times d^l}$ on a graph $\mathcal{G}$, where n denotes the number of nodes and d indicates the dimensionality of node features. The function $f(\cdot,\cdot)$ of GCN layer l takes the feature matrix H^l and a representative matrix A of the graph $\mathcal{G}$ structure as inputs, and updates the node features as $H^{l+1} \in \mathbb{R}^{n\times d^{l+1}}$. Every GCN layer can then be written as a non-linear function, accordingly:

$$H^{l+1} = f(H^l, A) \quad (2)$$

Following the layer-wise propagation rule of [13]:

$$f(H^l, A) = \sigma(\hat{A}H^l W^l) \quad (3)$$

where $W^l \in \mathbb{R}^{d^l \times d^{l+1}}$ is a weight matrix to be learned and $\sigma(\cdot)$ is a non-linear activation function like the RELU. $\hat{A}$ denotes the normalised version of A.

In the present work, stacked GCN layers were used to learn the inter-relationships of labels. The input of the first GCN layer is the $X \in \mathbb{R}^{C \times d}$ with C denoting the number of categories and d denoting the dimensionality of the one-hot label representation. For the last layer, the output matrix is $Z \in \mathbb{R}^{C \times D}$, where D is the dimensionality of the ECG representation.

Eq. 3 shows that the layer-wise propagation of GCN is based on a normalised matrix $\hat{A}$ which describes the graph structure in a matrix form. To construct matrix $\hat{A}$, the co-occurrence patterns of labels was mined, and then a correlation matrix between labels was defined. The label correlation dependency was modelled as the conditional probability. As shown in Fig. 2, each entry (x,y) of the matrix represents the probability of the occurrence of label x (along rows) when label y (along column) appears. To calculate the conditional probability in each entry (x,y), the occurrence of label pairs is counted and divided by the occurrence of label y in the training set, $i.e.$, $P(x|y) = \frac{P(x,y)}{P(y)}$.

Binary cross-entropy is widely used for multi-label classification problems, however it evaluates each label independently without considering label correlation [14]. Here, the output of our GCN network is introduced and a class-aware binary cross-entropy loss is defined as follows:

$$\mathcal{L} = \sum_{c=1}^{C} y^c \log(\frac{1}{1+e^{-F \cdot Z}}) + (1-y^c) \log(1-\frac{1}{1+e^{-F \cdot Z}}) \tag{4}$$

where $y^c = \{0, 1\}$ denotes the corresponding target binary value of label c.

2.3. Training

All 12-ECG recordings were firstly resampled to a 500 Hz sampling rate. The ECG recordings which are shorter than 18 seconds were zero-padded. Our network took this signal as input and output one prediction every 512 samples. The class-aware binary cross-entropy loss between the predictions and the labels from the training set was applied to optimise our network. Our network was trained for 100 epochs using Adam stochastic gradient descent (SGD) optimiser with random initialisation of the weights. The batch size and the learning rate were set to 32 and 0.001 respectively. The learning rate was reduced by a factor of 10 when the validation loss stopped improving for two consecutive epochs.

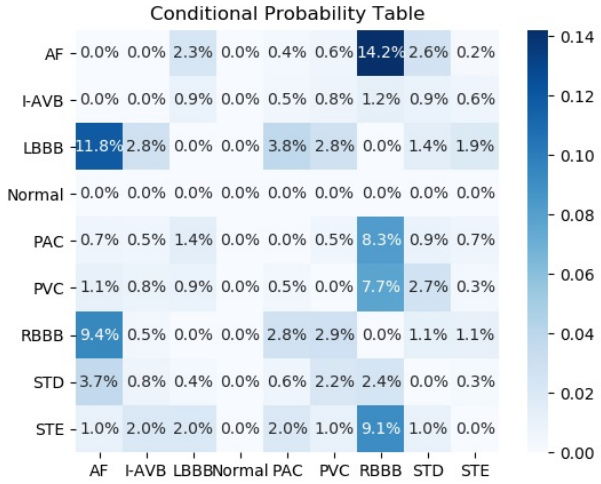

Figure 2. Conditional probability table between nine labels.

3. Experiment

3.1. Dataset

The dataset was the one provided by the PhysioNet/CinC Challenge 2020. This dataset was from multiple sources:China Physiological Signal Challenge in 2018 (10330 recordings), St. Petersburg INCART 12-lead Arrhythmia Database (75 recordings), Physikalisch Technische Bundesanstalt (22386 recordings) and Georgia 12-Lead ECG Challenge Database (10344 recordings). Each ECG recording has one or more labels from different type of abnormalities in SNOMED-CT codes.

3.2. Evaluation Metric

The first evaluation metric in the competition is designed to award full credit to correct diagnoses and partial credit to misdiagnoses. It is calculated as follows:

$$s = \Sigma_{ij} w_{ij} a_{ij} \tag{5}$$

where a_{ij} is the number of recordings in a database that were classified to class $i \in C$ but actually belong to class $j \in C$. All a_{ij} construct a multi-class confusion matrix. A weight matrix $W = [w_{ij}]$ was pre-defined based on the similarity of treatments or differences in risks.

Macro-F1 was also adopted as our second evaluation metric, which is calculated by averaging the F1 values over all the classes, as shown below:

$$Macro\text{-}F_1 = \frac{1}{C} \sum_{i=1}^{C} \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{6}$$

where C represents the number of classes ($C = 25$ in our case), TP, FP and FN represent the numbers of true positive, false positive and false negative samples respectively.

4. Results

In the present work, four controlled experiments were conducted to verify the effectiveness of every component in our proposed framework including Residual Blocks, bi-directional GRU and GCN. For each experiment, 5-fold cross-validation was performed on training set. The incremental development of our approach is illustrated in Table 1. An one dimension CNN (1DCNN) approach without residual blocks was firstly tested, where the Challenge Metric and F1 score were 0.503 and 0.481. Then we augmented the CNN architecture with the residual blocks and increased the Challenge Metric and the F1 score to 0.554 and 0.526. Afterwards, we added the bi-directional GCN and improved the Challenge Metric and the F1 score further to 0.582 and 0.564. For above three benchmarks, typically binary cross-entropy loss is applied to train the whole network. Finally, we introduced GCN and class-aware binary cross-entropy loss and achieved final Challenge Metric (0.627) and F1 (0.603) score. In the PhysioNet/Computing in Cardiology Challenge 2020, our team, Leicester-Fox, reached a challenge validation score of 0.395, and full test score of -0.012, placing us 34 out of 41 in the official ranking.

Table 1. Challenge Metric and F1 score of incremental development in our approach.

Methods	Challenge Metric	F_1 score
1DCNN	0.503	0.481
Res-Blocks	0.554	0.526
Res-Blocks + Bi-GRU	0.582	0.564
Res-Blocks+Bi-GRU+GCN	0.627	0.603

5. Conclusion

In this paper, we developed a deep neural network architecture for multi-label classification of cardiac abnormalities from 12-lead ECGs. The network contains two modules: the ECG representation network for learning high-level feature representation of ECG recordings and the GCN for capturing the inter-class relationships. Empirical evaluations demonstrated the effectiveness and efficacy of our architecture.

References

[1] Alday EAP, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad AB, Elola A, Seyedi S, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol. Meas. 2020 (Under review).

[2] Wilson FN, Kossmann CE, Burch GE, Goldberger E, Graybiel A, Hecht HH, Johnston FD, Lepeschkin E, Myers GB. Recommendations for standardization of electrocardiographic and vectorcardiographic leads. Circulation 1954;10(4):564–573.

[3] Schläpfer J, Wellens HJ. Computer-interpreted electrocardiograms: benefits and limitations. Journal of the American College of Cardiology 2017;70(9):1183–1192.

[4] Vijayavanan M, Rathikarani V, Dhanalakshmi P. Automatic classification of ecg signal for heart disease diagnosis using morphological features. International Journal of Computer Science Engineering Technology 2014;5(4):449–455.

[5] Korürek M, Doğan B. Ecg beat classification using particle swarm optimization and radial basis function neural network. Expert systems with Applications 2010; 37(12):7563–7569.

[6] Park K, Cho B, Lee D, Song S, Lee J, Chee Y, Kim IY, Kim S. Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. In 2008 Computers in Cardiology. IEEE, 2008; 229–232.

[7] Andreao RV, Dorizzi B, Boudy J. Ecg signal analysis through hidden markov models. IEEE Transactions on Biomedical engineering 2006;53(8):1541–1549.

[8] Kropf M, Hayn D, Schreier G. Ecg classification based on time and frequency domain features using random forests. In 2017 Computing in Cardiology (CinC). IEEE, 2017; 1–4.

[9] Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. Journal of electrocardiology 2007;40(5):385–390.

[10] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine 2019; 25(1):65.

[11] Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Wagner Jr M, et al. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications 2020;11(1):1–9.

[12] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In Advances in neural information processing systems. 2016; 3844–3852.

[13] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv160902907 2016;.

[14] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine learning 2011;85(3):333.

Address for correspondence:

Zheheng Jiang
Informatics, University Rd, Leicester, UK LE1 7RH
zj53@leicester.ac.uk