2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Improving Automatic Detection of ECG Abnormality with Less Manual Annotations using Siamese Network*

Fan Yang[1], Guijin Wang[1(✉)], Chuankai Luo[1], Zijian Ding[1]

Abstract— **Electrocardiography is a very common, non-invasive diagnostic procedure and its interpretation is increasingly supported by automatic interpretation algorithms. Recently many works also focused on the design of automatic ECG abnormality detection algorithms. However, clinical electrocardiogram datasets often suffer from their heavy needs for expert annotations, which are often expensive and hard to obtain. In this work, we proposed a weakly supervised pretraining method based on the Siamese neural network, which utilizes the original diagnostic information written by physicians to produce useful feature representations of the ECG signal which improves performance of ECG abnormality detection algorithms with fewer expert annotations. The experiment showed that with the proposed weekly supervised pretraining, the performance of ECG abnormality detection algorithms that was trained with only 1/8 annotated ECG data outperforms classical models that was trained with fully annotated ECG data, which implies a large proportion of annotation resource could be saved. The proposed technique could be easily extended to other tasks beside abnormality detection provided that the text similarity metric is specifically designed for the given task.**

Clinical Relevance—**This work proposes a novel framework for the automatic detection of cardiovascular disease based on electrocardiogram.**

I. INTRODUCTION

Automatic annotation algorithm of electrocardiogram (ECG) has received wide attention because of its vital role in clinical diagnosis of various cardiac diseases[1]. As an external measure result of myocardial electrophysiological activity, it reflects the electrical depolarization and repolarization patterns of the heart[2]. Arrhythmia is understood as disturbance in the rate, regularity, site of origin or conduction of the electrical impulses through the heart. The diagnosis of arrhythmia by ECG has attracted great attention from cardiologists[3,4,5,6,7,8,9,10,11,12,13, 14].

In recent years, many machine learning methods have been developed in ECG characteristic points detection. Saini et al.[15] proposed a K-Nearest Neighbor classification approach for ECG recognition. However, the K-NN method suffers from the curse of dimensionality when the feature dimension is high, and the trained K-NN classifier model is memory consuming since it needs to store all of the training data. Bayesian method was proposed[16,17] based on partially collapsed Gibbs sampler. By exploiting the strong local dependency of ECG signals, the method showed relatively high detection rate on QT database. Gao et al[18] proposed randomly selected signal pair difference (RSSPD) feature extracted from time domain signals and used a random forest classifier and some post processing to generate final results. Ming Chen et al[24] proposed end-to-end deep learning based ECG characteristic point detection algorithm, where region aggregation module was designed to replace the simple fully connected layers which usually play the role of regressor, which performed well on QT database[19].

Machine learning based ECG abnormality detection usually requires a large amount of data and high-quality annotations during the training. However, the clinical ECG signal annotations are expensive and difficult to obtain, and the precious time of cardiovascular physicians should be put into the diagnosis and treatment of patients. Most of the previous work focused on the design of efficient algorithms and models to improve the performance of the ECG signal abnormality detection algorithm, while failed to take consideration into the cost of model training. In order to facilitate the implementation and deployment of the ECG signal abnormality monitoring algorithm, we should explore cheaper training methods with less data, or at least, fewer manual annotations, to efficiently train the model to diagnose ECG abnormality.

In this work, we proposed a weakly supervised pretraining method based on the Siamese neural network, which utilizes the original diagnostic information of the ECG signal to calculate the semantic similarity between the ECG signals without the ECG signal category labeling, and then took the semantic similarity as the label during the training of the Siamese neural network, which extracts the representation of the ECG signals. After that, a Light GBM model was utilized to perform supervised training on a small fraction of manually annotated ECG data, obtaining a classifier on the abnormality of the ECG signal. Experiments show that such a design significantly reduced the need for manual labeling of ECG diagnoses during the training process, and allowed the ECG abnormality detection model to use less manual labeling information to obtain better classification performance on the PTB-XL dataset.

II. OUR APPROACH

A. Siamese Network

Siamese Networks[22] are widely used in the field of computer vision and natural language processing tasks to enhance the interpretability of neural network and to improve the classification accuracy on imbalanced datasets. A base network structure was selected, for example, one-dimensional residual network with 34 layers (i.e. ResNet34), as the component of the Siamese network. The final fully connected

* This work is partially funded by The National Key Research and Development Program of China (2017YFB1401804) and The Medicine-Engineering Innovation Support Program of Tsinghua University (IDS-MSP-2019003).

[1] Department of Electronic Engineering, Tsinghua University, Beijing, China
✉ E-mail address: wangguijin@tsinghua.edu.cn

layer of the base network was removed to output features generated by residual blocks rather than the final predictions. In the Siamese network, the base network was duplicated as two copies with same initial weight values, and the optimizer simultaneously update these two copies to ensure they keep same weight during the training process. The features produced by two copies are then concatenated and passed into a fully connected layer followed by a sigmoid function to produce a prediction between 0 and 1. The forward propagation process is shown as Figure 1.

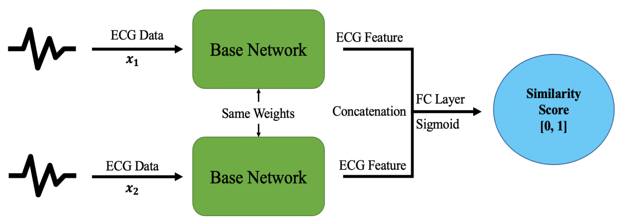

Figure 1: Forward propagation of proposed Siamese network

B. Semantic Similarity

The original diagnosis reports produced by doctors were contained in the PTB-XL dataset, which were written in Swedish and usually implied diagnosis conclusion for each patient, as illustrated in table 1. The original diagnosis reports were preprocessed at the text level to obtain the similarity between the diagnose report of each ECG signal, then served as labels during the training of the proposed Siamese neural network. The calculation of similarity is carried out using the *fuzzywuzzy* toolkit, which uses Levenshtein distance to define the similarity between two sentences and can be easily utilized on diagnosis reports written in other natural languages.

TABLE I. EXAMPLE OF PTB-XL DIAGNOSIS REPORTS

ECG ID	Diagnosis Report
8727	Sinusrytm vänster kammarhypertrofi.
9123	Sinus arrhythmie sonst normales ekg
13447	Sinusrytm ospecifikt skänkelblock avvikande qrs(t) förlopp inferior infarkt ålder ej bestämmbar.
13698	Sinusrhythmus qrs(t) abnormal septaler infarkt nicht auszuschliessen
…	…

C. ODENet

The ODENet was originally proposed by Chen et al[23], achieving state-of-the-art performance on MNIST[25] dataset. The ODENet contains several blocks named ODE Blocks, as shown in Figure 2, which retains feature dimensions and could be replicated and reused for multiple time to reduce the spatial complexity of model. The forward propagation and back propagation of the ODENet differs from traditional neural networks. During the forward propagation, each ODE blocks were reused for multiple times to simulate the process of solving an ODE. While during the backward propagation, the gradient of the output of each ODE Block with respect to the input of each ODE Block was calculated by adjoint method rather than the chain rule, which avoids the gradient vanishing problem when the ODE Block is deep. We will show that the ODENet alone achieve state-of-the-art performance on the PTB-XL dataset, and the Siamese neural network consisted of ODENets also achieve state-of-the-art performance on abnormality detection task.

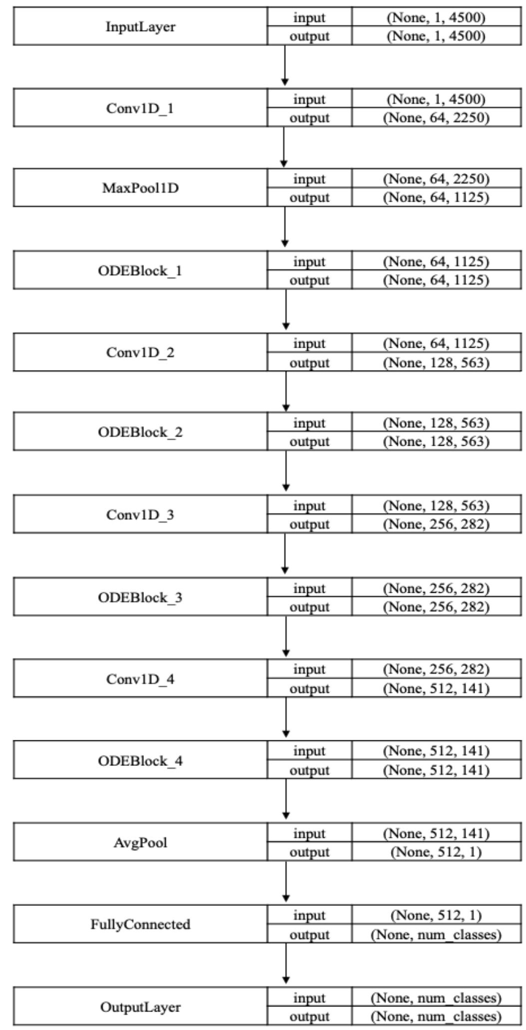

Figure 2: Structure of the proposed ODE Net

III. EXPERIMENTS AND DISCUSSIONS

We use the Siamese network and ODENet proposed above on the diagnostic classification tasks on the PTB-XL dataset. The classification performance showed that the ODENet alone can achieve state-of-the-art ECG abnormality detection performance with much fewer network parameters compared with neural networks with other architectures, and the Siamese neural network can achieve state-of-the-art ECG abnormality detection performance with much less manual annotation of training ECG signals, which demonstrates the Siamese network learned good features representations for abnormality classification.

A. PTB-XL Dataset

The PTB-XL dataset[20] comprises 21837 clinical 12-lead ECG records of 10 seconds length from 18885 patients, where 52% were male and 48% were female. The ECG statements used for annotation are conform to the SCP-ECG standard and were assigned to three non-mutually exclusive categories diag. (short for diagnostic), form and rhythm. In total, there are 71 different statements, which decompose into 44 diagnostic, 12 rhythm and 19 form statements. Note that there are 4 form statements that are also assigned to the set of diagnostic ECG statements. For diagnostic statements also a hierarchical organization into five coarse super-classes (NORM: normal ECG, CD: conduction disturbance, MI: myocardial infarction, HYP: hypertrophy and STTC: ST/T changes) and 24 sub-classes is provided, see Figure 3. For further details on the dataset and the annotation scheme, we refer the readers to the original publication[20]. We also refer the readers to the publication[21] with benchmarks and insights of PTB-XL for a more comprehensive understanding of the dataset.

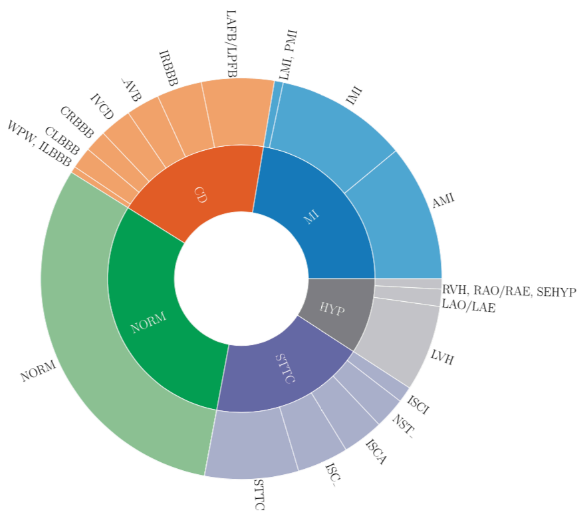

Figure 3: Sample distribution of super-classes and sub-classes in PTB-XL dataset

B. Experiment Setting

Before the training process, a median filter was applied on the PTB-XL ECG data to remove baseline drift, and a Daubechies wavelet filtering was applied to remove high frequency noise. The sampling frequency for the PTB-XL dataset is 100 Hz, and we split each ECG record into 1000 sampling points (i.e. 10 seconds). The training set, validation set and test set are divided according to the 8:1:1 ratio recommended by the publisher of the PTB-XL dataset. We run the training on a single GeForce RTX GPU with maximum power of 250W. During the testing process, we use F1-macro and AUC as the indicator to evaluate the performance of the trained ODENet and Siamese network on the testing set. We evaluate our algorithms on both 12-leads dataset and single lead (I-lead) dataset to address their performance on clinical ECGs and dynamic ECGs.

C. Experiment Results - ODENet

The evaluation results of ODENet's performance on both 12-leads and single lead PTB-XL dataset are shown as Table 2 and Table 3. We may conclude that the overall performance of ODENet is the same as ResNet34 and XResNet101, which are the current best baseline of ECG classification tasks. Especially for 12-leads abnormality detection task, ODENet outperforms both ResNet34 and XResNet101. The reason might be that the abnormality features were more likely to be captured by Siamese neural network than other categories.

TABLE II. ODENET'S PERFORMANCE ON 12-LEADS DATASET

	ResNet34		XResNet101		ODENet	
	F1	**AUC**	**F1**	**AUC**	**F1**	**AUC**
NORM	**0.8268**	**0.8952**	**0.8043**	**0.9008**	**0.8516**	**0.9166**
MI	0.7940	0.8908	0.7940	0.8908	0.7485	0.8707
STTC	0.8116	0.9051	0.8123	0.9003	0.8040	0.8960
CD	0.8555	0.9322	0.8681	0.9341	0.8594	0.9296
HYP	0.8264	0.8904	0.8370	0.9298	0.8265	0.9221
Average	**0.8229**	**0.9027**	**0.8231**	**0.9112**	**0.8180**	**0.9070**

TABLE III. ODENET'S PERFORMANCE ON 1-LEAD DATASET

	ResNet34		XResNet101		ODENet	
	F1	**AUC**	**F1**	**AUC**	**F1**	**AUC**
NORM	**0.7546**	**0.8125**	**0.7599**	**0.8199**	**0.7446**	**0.7992**
MI	0.6848	0.7906	0.6561	0.7629	0.7020	0.8073
STTC	0.7173	0.7996	0.7167	0.7904	0.7054	0.7899
CD	0.8095	0.8804	0.8118	0.8861	0.8147	0.8877
HYP	0.7694	0.8632	0.7722	0.8672	0.7655	0.8682
Average	**0.7471**	**0.8297**	**0.7433**	**0.8253**	**0.7464**	**0.8305**

D. Experiment Results – Siamese Network

The evaluation results of the proposed Siamese neural network's performance compared with single ODENet on both 12-leads and single lead PTB-XL dataset with different annotation proportions on abnormality detection task are shown as Table 4 and Table 5. We may conclude that for single ODENet, the classification performance on abnormality detection task significantly suffers from the lack of data annotation. However, for the Siamese neural network, the weakly supervised pretraining allows the base network to learn useful representations for classification tasks only by original diagnosis reports without manual annotation by human experts, which improves the classification performance especially when there are only a small fractions of ECG data annotated (e.g. 1/8). For example, for 12-lead PTB-XL dataset, the performance of the Siamese neural network with only 1/8 ECG training data annotated outperforms the performance of single ODENet with 5/8 ECG training data annotated, and is almost the same as the performance of single ODENet with all ECG training data annotated. For I-lead PTB-XL dataset, the performance of the Siamese neural network with only 1/8 ECG training data annotated outperforms the performance of single ODENet with all ECG training data annotated, which implies a large proportion of annotation resource could be saved.

TABLE IV. SIAMESE'S PERFORMANCE ON 12-LEADS DATASET

Annotation	**ODENet**		**ODENet Siamese**	
Ratio	**F1**	**AUC**	**F1**	**AUC**
1/8	0.7808	0.8498	**0.8254**	**0.9111**
3/8	0.7979	0.8695	0.8338	0.9122
5/8	**0.8136**	**0.8893**	0.8307	0.9151
8/8	**0.8516**	**0.9166**	0.8327	0.9146

TABLE V. SIAMESE'S PERFORMANCE ON 1-LEAD DATASET

Annotation Ratio	**ODENet**		**ODENet Siamese**	
	F1	**AUC**	**F1**	**AUC**
1/8	0.6838	0.7209	**0.7746**	**0.8499**
3/8	0.7082	0.7426	0.7667	0.8439
5/8	0.7445	0.8008	0.7791	0.8543
8/8	**0.7446**	**0.7992**	0.7881	0.8528

IV. CONCLUSION

In this work, we tackled the well-known problem in the field of intelligent medicine that clinical electrocardiogram datasets often suffer from their heavy needs for expert annotations, which are often expensive and hard to obtain. To address this problem, we proposed a weakly supervised pretraining method based on the Siamese neural network, which utilizes the original diagnostic information of the ECG signal to calculate the semantic similarity between the ECG signals without the ECG signal category labeling, and then took the semantic similarity as the label during the training of the Siamese neural network, which extracts the representation of the ECG signals. Experiments show that such a design significantly reduced the need for manual labeling of ECG diagnoses during the training process, and allowed the ECG abnormality detection model to use less manual labeling information to obtain better classification performance on ECG dataset, which implies a large proportion of annotation resource could be saved. The proposed technique could be easily extended to other tasks beside abnormality detection provided that the text similarity metric is specifically designed for the given task.

REFERENCES

[1] A. Gacek and W. Pedrycz, ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence. Springer Science & Business Media, 2011.

[2] G.J. Taylor, 150 Practice ECGs: Interpretation and Review, Wiley, 3rd ed., 2008.

[3] P. de Chazal, R.B. Reilly, A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2535–2543, doi:10.1109/TBME.2006.883802.

[4] O.T. Inan, L. Giovangrandi, G.T.A. Kovacs, Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2507–2515, doi:10.1109/TBME.2006.880879.

[5] T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, IEEE Trans. Biomed. Eng. 56 (5) (2009) 1415–1426, doi:10.1109/TBME.2009.2013934.

[6] W. Jiang, S.G. Kong, Block-based neural networks for personalized ECG signal classification, IEEE Trans. Biomed. Eng. 18 (6) (2007) 1750–1761, doi:10.1109/TNN.2007.900239.

[7] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-d convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (3) (2016) 664–675, doi:10.1109/TBME.2015.2468589.

[8] H. Li, D. Yuan, X. Ma, Genetic algorithm for the optimization of features and neural networks in ECG signals classification, Sci. Rep. 7 (2017), doi:10.1038/srep41011.

[9] S. Osowski, L.T. Hoai, T. Markiewicz, Support vector machine-based expert system for reliable heartbeat recognition, IEEE Trans. Biomed. Eng. 51 (4) (2004) 582–589, doi:10.1109/TBME.2004.824138.

[10] M.M.A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inf. Sci. (Ny) 345 (1) (2016) 340–354, doi:10.1016/j.ins.2016.01.082.

[11] E.J. da S.Luz, T.M. Nunes, V.H.C. de Albuquerque, J.P. Papa, D. Menotti, ECG arrhythmia classification based on optimum-path forest, Expert Syst. Appl. 40 (9) (2013) 3561–3573, doi:10.1016/j.eswa.2012.12.063.

[12] J.S. Wang, W.C. Chiang, Y.L. Hsu, Y.T.C. Yang, ECG arrhythmia classification using a probabilistic neural network with a feature reduction method, Neurocomputing 116 (10) (2013) 38–45, doi:10.1016/j.neucom.2011.10.045.

[13] J. Wiens, J.V. Guttag, Active learning applied to patient-adaptive heartbeat classification, Adv. Neural Inf. Process Syst. (2010) 2442–2450.

[14] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, H. Yang, Patient-specific ECG classification based on recurrent neural networks and clustering technique, in: Proceedings of the 2017 Thirteenth IASTED International Conference on Biomedical Engineering (BioMed), 2017, pp. 63–67, doi:10.2316/P.2017.852-029.

[15] I. Saini, D. Singh, and A. Khosla, "Delineation of ecg wave components using k-nearest neighbor (knn) algorithm: Ecg wave delineation using knn," in Information Technology: New Generations (ITNG), 2013 Tenth International Conference on. IEEE, 2013, pp. 712–717.

[16] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P-and t-wave delineation in ecg signals using a bayesian approach and a partially collapsed gibbs sampler," IEEE transactions on biomedical engineering, vol. 57, no. 12, pp. 2840–2849, 2010.

[17] C. Lin, G. Kail, A. Giremus, C. Mailhes, J.-Y. Tourneret, and F. Hlawatsch, "Sequential beat-to-beat p and t wave delineation and waveform estimation in ecg signals: Block gibbs sampler and marginalized particle filter," Signal Processing, vol. 104, pp. 174–187, 2014.

[18] P. Gao, J. Zhao, G. Wang, and H. Guo, "Real time ecg characteristic point detection with randomly selected signal pair difference (rsspd) feature and random forest classifier," in Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE, 2016, pp. 732–735.

[19] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in Computers in Cardiology 1997. IEEE, 1997, pp. 673–676.

[20] P.Wagner,N.Strodthoff,R.-D.Bousseljot,W.Samek,andT.Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," Phy- sioNet, https://doi.org/10.13026/6sec-a640, 2020.

[21] Strodthoff N , Wagner P , Schaeffter T , et al. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL[J]. 2020.

[22] Bertinetto L , Valmadre J , Henriques J F , et al. Fully-Convolutional Siamese Networks for Object Tracking[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[23] Chen R , Rubanova Y , Bettencourt J , et al. Neural Ordinary Differential Equations[J]. 2018.

[24] Chen M , Wang G , Ding Z , et al. Unsupervised Domain Adaptation for ECG Arrhythmia Classification[C]// 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. IEEE, 2020.

[25] Lecun Y , Cortes C . The mnist database of handwritten digits[J]. http://www.research.att.com/yann/ocr/mnist/, 2010.