Deep discriminative domain generalization with adversarial feature learning for classifying ECG signals

Deep Discriminative Domain Generalization with Adversarial Feature Learning for Classifying ECG Signals

Zuogang Shang[1],Zhibin Zhao[1],Hui Fang[2],Samuel Relton[3], Darcy Murphy[4], Zoe Hancox[3], Ruqiang Yan[1],David Wong[4]

[1] Xi'an Jiaotong University, Xi'an, China
[2] Loughborough University, Loughborough, UK
[3] University of Leeds, Leeds, UK
[4] University of Manchester, Manchester, UK

Abstract

Introduction: The goal of the 2021 PhysioNet/CinC challenge is diagnosing cardiac abnormalities from electrocardiograms (ECGs) and evaluating the diagnostic potential of reduced-lead ECGs. We describe the whole model created by the team "AI_Healthcare" for this goal.

Methods: ECGs were downsampled to 300 Hz and filtered by wavelet. Then ECGs we randomly clipped or zero-padded to 4,096 samples. To have a better representative learning ability, a modified ResNet with larger kernel sizes was used. Multi-source adversarial feature learning was used to learn domain-invariant and discriminative representations with a special gradient reversal layer (GRL). The performance with and without the domain generation methods was compared.

Results: We achieved a challenge score of 0.66, 0.64, 0.65, 0.65, 0.62 on the validation data. We ranked 8th, 7th, 6th, 6th, and 12th for 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs, respectively. Testing showed that domain generation improved metric scores on the unseen domain.

Conclusion: Generalized representations perform well for "unseen" data. It is a general method for other models to improve generalization performance by learning a domain-invariant feature representation.

1. Introduction

In this paper we describe the deep learning model developed for cardiac abnormality classification using 12-lead, 6-lead, 4-lead, 3-lead and 2-lead electrocardiogram (ECG) signals with varying sample lengths and frequencies.

Typically 12-lead ECGs are used clinically to diagnose cardiac abnormalities by measuring the electrical activity of the heart. Reduced-lead ECGs are also being explored for their diagnostic potential to reduce recording time, lessen expense and improve ease of use required in clinical settings [1].

Clinical experts of ECG interpretation are often required for cardiac abnormality diagnosis, which often takes a significant time from the cardiologist to view and document the findings. If a cardiologist is unavailable another non-specialist clinician may be required to make a diagnosis. Additionally, without 12-lead ECG recordings it is more difficult to observe the ST segment of the ECG wave, which could be vital for diagnosis of a subset of cardiac abnormalities such as ischemia [2].

Diagnostic tools using computational methods may reduce resource demand for cardiac abnormality diagnosis [3]. ECG classification using deep learning models, such as the one described in this paper, may have the ability to automatically diagnose a range of cardiac abnormalities without requiring all 12-leads.

We set about this task by building on the work of the previous deep neural network architecture [4], incorporating domain generalisation through adversarial feature learning.

2. Methods

Our goal was to create a model to evaluate the diagnostic potential for reduced-lead ECGs. We used a modified ResNet with a Squeeze-and-Excitation (SE) attention block to extract deep features. Combined with hand-crafted features, a multi-source adversarial network was trained to learn domain-invariant representations which were discriminative for the main task of diagnosing cardiac abnormalities.

2.1. Data Pre-processing

Compared with last year's challenge [5], more datasets were available for training and kept for testing. 88,253 recordings from six datasets were provided for the

2021 PhysioNet/CinC challenge; CPSC, INCART, PTB, G12EC, Chapman-Shaoxing, and Ningbo. 42,902 recordings were kept as test data. The test recordings were from CPSC, G12EC, an undisclosed American database, and a new undescribed database [1].

The training dataset contained 133 diagnoses (classes) in total. 30 classes were considered for scoring the challenge, with 4 pairs being considered equivalent, making this effectively a 26 class problem. Some recordings had multiple diagnoses, and we returned a blank label for a recording which had no scored diagnosis.

Frequency and duration varied between recordings. All ECGs were resampled to 300 Hz for input to the deep model. During the training phase, we chose a signal length of 4096 samples (13.653s). Shorter signals were randomly zero-padded and longer signals were randomly clipped. The 2-lead signals were randomly clipped into two separate channels.

To reduce the influence of environment noise, we employed wavelet denoising as in [6]. As the frequency was 300 Hz, ECGs were decomposed into 9 levels with Daubechies D_6 ('db6') wavelet. Useful information for cardiac abnormalities is available in the frequency range of 0.5-40 Hz [7]. The first approximation sub-band consisted of frequency range of 0-0.293 Hz which was mainly the baseline wander. The first detail sub-band consisted of the frequency range of 75-150 Hz which contained little relevant information. We replaced the first approximation sub-band and the first detail sub-band with zeros. The other detail sub-bands were used to reconstruct the signal.

Age, gender, and special Heart Rate Variability (HRV) features were used as extra features and were concatenated with deep features. Unknown values of age and gender were masked and set to 0.

As all reduced-lead ECGs contained lead I and lead II, we chose these two leads to calculate HRV features. First, R peak locations were extracted in lead I and lead II using the EngZee QRS detector [8]. Then, these peaks were used to derive: the standard deviation of R-peak, or normal-to-normal, (SDNN) intervals, root mean square of successive R-peak differences (RMSSD), the standard deviation of the successive differences (SDSD) between adjacent R peak (NN) intervals, the proportion of NNs that are greater than 20 ms (NN20) divided by total number of R-peak intervals (PNN20), and heart-rate (HR). For normalization, SDSD was divided by 1000 and HR was divided by 100. If there were fewer than 5 detected R peaks per example, we assumed that there was mistake in R peak detection and HRV features were defined as unknown for this recording, in which case it was masked by setting HRV features to zeros. As shown in Fig 1, age, gender, and HRV features were encoded to a total of 17 feature values.
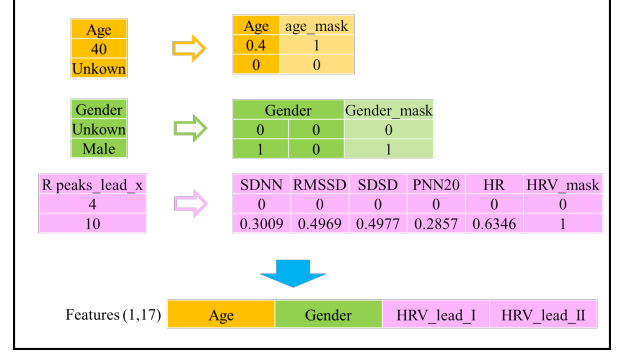

Figure 1. Input features of age, gender, and HRV.

2.2. Model Description

A domain generation model was designed to extract discriminative domain-invariant features from the input signals and extra features. This model classified the ECG recordings into 26 classes. The model structure is illustrated in Fig 2.
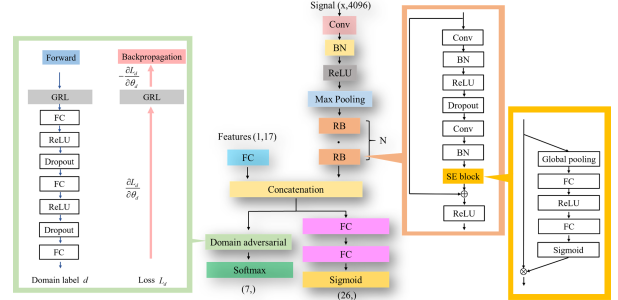

Figure 2. Architecture of the proposed model.

2.2.1. ResNet Feature Extraction

The modified ResNet model with an adaptive input channel set was used for all of the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead input signals. The modified ResNet model from [4] consists of one convolution layer with a wide kernel and 8 residual blocks (RBs).

A wider kernel in the first layer has been shown to perform better in time sequence classification tasks [9]. We employ a convolution kernel size of 15 in the first layer followed by a batch normalisation (BN) layer and a rectified linear unit (ReLU). 64 kernels are used in the first convolution layer.

The RB consists of two convolution layers. Between the 2 layers, BN and ReLU are used. A dropout layer is also inserted to alleviate overfitting with a dropout rate of 0.2. After the second convolution, a BN layer and a SE block

[10] are used, followed by a residual connection from RB input and a ReLU layer. A wide convolution kernel with kernel size of 7 is employed in the RB. The number of kernels for the RB are 64, 64, 128, 128, 256, 256, 512, and 512. The feature dimension is halved after the third, fifth, and seventh RB. The SE block acts to adaptively recalibrates channel-wise feature response and calculates channel importance by explicitly modelling the independencies between channels. The SE block contains a global average pooling layer, a bottleneck with two fully connected (FC) layers around a ReLU layer, and a sigmoid layer. The reduction between the two FC layers is 16.

8 RBs are used to enlarge model receptive field and improve feature extraction ability. The residual connection confirms the training process stability [11].

2.2.2. Domain Classifier

As the datasets are from several hospitals with different environments and demographics, there are domain specific features in each dataset which influence feature extraction. Fewer leads mean reduced information for the model to use. It is easier for the deep model to learn domain representation when leads are reduced, especially for two leads. Domain-invariant representations are expected to improve the model's ability to generalize to unseen datasets as it is more likely that the features will be applicable to the new dataset.

Data from different domains (datasets) have a shift in distributions and representations [12]. We envisage that the final classification decisions should be made based on representations that are both discriminative for the main task and invariant to the domain changes. A discriminative domain-invariant representation requires mapping a domain-variant representation into a similar representation in different domains. We divided our training datasets into six domains by their recording file name and gave each ECG recording a domain label. The domain classifier consists of a simple three-layer bottleneck FC classifier and a GRL. The output feature dimensions for three FC layers are 1024, 1024, and 26. By minimizing the domain label prediction loss L_2, the domain classifier is optimized to learn domain features from input features. With the GRL, the input features from different domains are optimized to be similar to each other, and difficult to classify into the correct source domain. This is so the feature representations are mapped into a domain-invariant space.

2.2.3. Discriminative Classifier

SE-ResNet extracts deep features from the input ECGs. Using a FC layer, the encoded HRV features, age, and sex are mapped into extra features with a dimension of 34. The deep features and extra features are concatenated to a total

dimension of 546. Backpropagation, domain classifier and GRL are used to alter features to become domain-invariant. The final multi-label classification decisions are made by two FC layers with a middle dimension of 256. The loss for the discriminative classifier is L_1.

2.2.4. Training Setup

The training error for multi-label classification was average binary cross entropy (BCE) loss L_1. For the adversarial domain classification task, the loss was cross entropy L_2. The final loss L is:

$$L = L_1 + \lambda L_2 \qquad (1)$$

In which the weight parameter, λ, was set empirically at 0.05. For training, we chose 0.0003 as the initial learning rate with the Adam optimiser. It was reduced tenfold in the 20th epoch. The model was trained for a total of 30 epochs with batch size of 64.

2.3. Model Evaluation

Thresholds for different classes should be different because of the class imbalance. For our multi-label task, the decision boundaries of the classifier were adjusted by changing the thresholds for each class. After training, we used the validation signals to search for the best thresholds for the models. The search consisted of two steps: (1) Thresholds are initialised to be the same for all classes and then searched in the range [0,1] with a step 0.1 to get an approximate threshold; (2) Adjust approximate threshold for each class by searching with in steps of 0.01 when all other thresholds are fixed.

Validation signals shorter than 4,096 were zero-padded. Longer signals were segmented them into multiple patches. The overlap is $O = 256$ with a adaptive overlap for the last patch. The number of patches P can be calculated as follows:

$$P = ceil(\frac{L - 4096}{4096 - O}) + 1 \qquad (2)$$

3. Results

To evaluate our domain generation model we trained the model on three of the datasets and withheld the Ningbo dataset to use as a test *unseen domain*. We used the CPSC 2018, PTB, and G12EC databases for training, which we term the *seen domain*. During the training process, we treated the three databases as three domains.

As a comparator, we trained a baseline model without domain adversarial features which we called the *normal model*. The results of five-fold cross-validation are shown in Table 1, and are calculated using the 2021 Challenge

scoring metric. In the seen domain the domain generation model had a worse performance because of the adversarial feature learning. The domain generation model obtained a better performance in the unseen domain compared to the normal model.

Table 1. Evaluation of the domain generation method. N1: normal model performance in the seen domain. D1: domain generation model performance in the seen domain. N2: normal model performance in the unseen domain. D2: domain generation model performance in the unseen domain.

Lead	12	6	4	3	2
N1 (seen)	0.75	0.71	0.73	0.73	0.71
D1 (seen)	0.72	0.68	0.69	0.69	0.68
N2 (unseen)	0.43	0.46	0.46	0.44	0.45
D2 (unseen)	0.44	0.49	0.48	0.48	0.49

During the official phase, we achieved our best challenge validation scores of 0.66, 0.64, 0.65, 0.65, and 0.62 for 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead respectively.

4. Discussion and Conclusions

The proposed deep learning model has the ability to classify 26 cardiac abnormalities using 12-lead, 6-lead, 4-lead, 3-lead or 2-lead ECGs.

In the prior PhysioNet competition we attained a 5-fold cross-validation metric score of 0.684 on 12-lead ECG data alone [4]. However, the current model has greater potential for application and generalizability thanks to its ability to extract important features from any of the 12-lead or reduced-lead ECGs, alongside feature generation from unique HRVs, age and gender.

We have demonstrated that the modified ResNet feature extraction combined with a domain and discriminative classifier can achieve good ECG classification scoring. Our model showed consistent performance between the different lead sets. This model showed good performance on the validation data, indicating its potential for practical clincal use. Although the domain generation model showed improvement on the unseen domain when compared to the normal model, both models did still have a noticeable drop in performance for the unseen domain. This may indicate that domain-invariant features are only part of the solution for making a more general model, as there may be genuine differences such as different diagnostic criteria which cannot be accounted for with domain-invariant features alone. Interestingly, the 12-lead ECG received the worst score in the unseen domain, this requires further investigation as this could suggest that reduced-lead ECGs are suitable for cardiac abnormality prediction with this proposed model.

References

[1] Reyna MA, Sadr N, Alday EAP, Gu A, Shah AJ, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, et al. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. Computing in Cardiology 2021;48:1–4.

[2] Drew BJ, Pelter MM, Brodnick DE, Yadav AV, Dempel D, Adams MG. Comparison of a new reduced lead set ecg with the standard ecg for diagnosing cardiac arrhythmias and myocardial ischemia. Journal of electrocardiology 2002; 35(4):13–21.

[3] Serhani MA, T El Kassabi H, Ismail H, Nujum Navaz A. Ecg monitoring systems: Review, architecture, processes, and key challenges. Sensors 2020;20(6):1796.

[4] Zhao Z, Fang H, Relton SD, Yan R, Liu Y, Li Z, Qin J, Wong DC. Adaptive lead weighted resnet trained with different duration signals for classifying 12-lead ecgs. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.

[5] Alday EAP, Gu A, Shah AJ, 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. Physiological measurement Nov. 2020;41(12):124003.

[6] Martis RJ, Acharya UR, Min LC. Ecg beat classification using pca, lda, ica and discrete wavelet transform. Biomedical Signal Processing and Control 2013;8(5):437–448.

[7] Sangaiah AK, Arumugam M, Bian GB. An intelligent learning approach for improving ecg signal classification and arrhythmia analysis. Artificial intelligence in medicine 2020;103:101788.

[8] Engelse WA, Zeelenberg C. A single scan algorithm for qrs-detection and feature extraction. Computers in cardiology 1979;6(1979):37–42.

[9] Zhang W, Peng G, Li C, Chen Y, Zhang Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 2017;17(2):425.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018; 7132–7141.

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

[12] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In International conference on machine learning. PMLR, 2015; 1180–1189.

Address for correspondence:

Zuogang Shang
International center of Machinery, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China.
administrator@stu.xjtu.edu.cn