---

**Proceedings Paper:**
Shang, Z, Zhao, Z, Fang, H et al. (5 more authors) (2021) Deep Discriminative Domain Generalization with Adversarial Feature Learning for Classifying ECG Signals. In: 2021 Computing in Cardiology (CinC). 2021 Computing in Cardiology (CinC), 13-15 Sep 2021, Brno, Czech Republic. IEEE . ISBN 9781665479165

https://doi.org/10.23919/CinC53138.2021.9662844

---

# 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 to classify cardiac abnormalities from ECGs and evaluate the diagnostic potential of reduced-lead ECGs. Here, we describe the classification model created by the team "AI_Healthcare".*

*Methods: ECGs were downsampled to 300 Hz and filtered by wavelet. ECGs were randomly clipped or zero-padded to 4,096 samples. We trained as SE-ResNet as a baseline classifier. We then modified it to classify both dataset and disease. We used a gradient reversal layer as part of an adversarial feature learning scheme to learn domain-invariant and discriminative representations. Performance with and without the domain generation methods was compared.*

*Results: In local validation on a held-out data set, our domain-invariant model achieved better challenge evaluation metric scores than the baseline SE-ResNet (12 lead: 0.43 vs 0.44, 2 lead: 0.45 vs 0.49). Only the baseline was tested on the hidden test set, achieving scores of 0.42, 0.42, 0.42, 0.44, and 0.38 on 12-leads, 6-leads, 4-leads, 3-leads, and 2-leads, respectively.*

*Conclusion: The domain generation method performed well on "unseen" data in local testing, suggesting that this method may help improve generalisation performance.*

## 1. Introduction

In this paper we describe a deep learning model developed to classify cardiac abnormality from 12-lead, 6-lead, 4-lead, 3-lead and 2-lead electrocardiogram (ECG) signals with varying sample lengths and frequencies.

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 and expense, and improve ease of use in clinical settings [1].

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, which could reduce resource demand. However, in the Physionet 2020 challenge [2], all models suffered from poorer performance on a hidden dataset from an undisclosed location.

We aim to address this issue by building on the work of the previous deep neural network architecture [3], incorporating domain generalisation through adversarial feature learning.

## 2. Methods

Our goal was to create a ECG classification model that learned domain-agnostic features, and that could also be applied to 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 useful domain-invariant features for the main task of diagnosing cardiac abnormalities. We expected that the domain-invariant representation would perform worse on the test data from the seen datasets. As the seen datasets have a frequency of 500Hz and the majority of the unseen test data is of a different frequency, we used a model without domain generalisation (baseline model) for test examples with a frequency of 500Hz.

### 2.1. Data Pre-processing

More datasets were available for training and kept for testing than in last year's challenge [1, 2]. 88,253 recordings were provided for the 2021 PhysioNet/CinC challenge. The datasets were CPSC [4], INCART [5], PTB [6], PTB-XL [7], G12EC, Chapman-Shaoxing [8], and Ningbo

[9].

30 cardiac condition classes were considered for scoring in the challenge. Four pairs of classes were considered equivalent, making this effectively a 26-class problem.

ECGs were resampled to 300 Hz for input to the deep model. During the training phase, we chose a signal length of 4096 samples. Shorter signals were randomly zero-padded and longer signals were randomly clipped.

To reduce unwanted noise, we employed wavelet denoising [10]. As the frequency was 300 Hz, ECGs were decomposed into 9 levels with Daubechies $D_6$ ('db6') wavelet. We replaced the first approximation sub-band (baseline wander) and the first detail sub-band (little relevant information) with zeros. The other detail sub-bands were used to reconstruct the signal.

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

HRV features were extracted from lead I and II. First, R peak locations were extracted using the EngZee QRS detector [11]. 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. All HRV features were set to zero if fewer than 5 R peaks were detected in an example. Age, gender, and HRV features were encoded to a total of 17 feature values.

## 2.2. Model Description

Our model was designed to extract discriminative domain-invariant features from the input signals and extra features. It then uses the features to classify the ECG recordings into 26 classes. It achieves this by multi-task learning of ECG abnormalities and domain, with a loss function that seeks to maximise domain loss, and minimise ECG abnormality loss. The model structure is illustrated in Fig 1.

### 2.2.1. ResNet Feature Extraction

The initial branch of the model is a modified ResNet model with an adaptive input channel. The modified ResNet model from [3] consists of one convolution layer with a wide kernel and 8 residual blocks (RBs).

A wide kernel in the first layer has been shown to perform better in time sequence classification tasks [12]. We employ a convolution kernel size of 15 in the first layer followed by batch normalisation (BN) 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 with dropout rate of 0.2 is also inserted to alleviate overfitting. After the second convolution, a BN layer and a SE block [13] are used, followed by a residual connection from RB input and a ReLU layer. A convolution kernel with 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 recalibrate channel-wise feature response and calculates channel importance by explicitly modelling the dependencies 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 [14].

After deep feature extraction, we concatenate the deep feature set with the encoded HRV, age, and gender features to a total dimension of 546.

The features are used for two tasks, domain classification and ECG abnormality classification.

### 2.2.2. Domain Classifier

Data from different domains (datasets) may have a shift in distributions and representations [15]. We envisage that the final classification decisions should be based on representations that are both discriminative for the main task (ECG abnormality classifcation) 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 seven 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 Gradient Reversal Layer (GRL). We label the loss for this branch as $L_2$.

By minimizing the domain label prediction loss $L_2$, the domain classifier is optimized to learn domain features from input features. The GRL means that the gradient for $L_2$ is reversed for the feature extraction part of the network, meaning that the feature extractor tries to maximise $L_2$. This leads to the feature extractor learning features which give the least domain information.

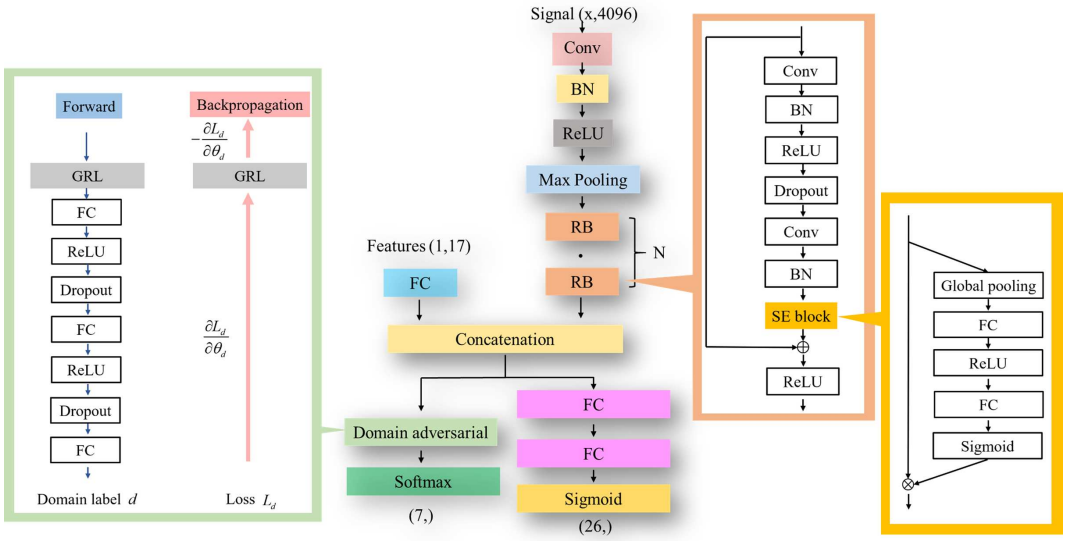


Figure 1. Architecture of the proposed model.

### 2.2.3. Discriminative Classifier

Multi-label ECG abnormality classifications are created from the 546-dimension features by using 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:

$$L = L_1 + \lambda L_2. \quad (1)$$

The weight parameter, $\lambda$, 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.

The baseline model was trained on CPSC and G12EC with the same parameters, using the $L_1$ loss only.

## 2.3. Model Evaluation

Thresholds for different classes should be different because of class imbalance. After training, we used the validation signals to search for the best thresholds for the models: (1) Thresholds were 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 into multiple patches. The overlap is $O = 256$ with a adaptive overlap for the last patch.

We expected that in testing most examples with a frequency of 500Hz would be from the same domain as the training data (CPSC and G12EC). We planned to use the baseline model for 500Hz examples and the domain-invariant model for other examples.

| Lead | 12 | 6 | 4 | 3 | 2 |
|---|---|---|---|---|---|
| B (seen domain) | 0.75 | 0.71 | 0.73 | 0.73 | 0.71 |
| D (seen domain) | 0.72 | 0.68 | 0.69 | 0.69 | 0.68 |
| B (unseen domain) | 0.43 | 0.46 | 0.46 | 0.44 | 0.45 |
| D (unseen domain) | 0.44 | 0.49 | 0.48 | 0.48 | 0.49 |

Table 1. Challenge metric scores for the baseline and domain invariant models for data in the seen and unseen domains. B: baseline model. D: domain invariant model.

## 3. Results

For local validation, we trained the model on N-1 of the N training data sets, reserving the Ningbo dataset as a local test. 5-fold cross-validation results are shown in Table 1. The domain generation model obtained a better performance in the unseen Ningbo dataset compared to the baseline model.

Due to a technical error, the domain-invariant model was

| Leads | Training | Validation | Test | Ranking |
|---|---|---|---|---|
| 12 | $0.721 \pm 0.001$ | 0.64 | N/A | none |
| 6 | $0.690 \pm 0.002$ | 0.63 | N/A | none |
| 4 | $0.708 \pm 0.002$ | 0.63 | N/A | none |
| 3 | $0.690 \pm 0.001$ | 0.64 | N/A | none |
| 2 | $0.685 \pm 0.001$ | 0.61 | N/A | none |

Table 2. Challenge scores for our domain-invariant model (team AI_Healthcare) using 5-fold cross validation on the public training set and repeated scoring on the hidden validation set. We did not receive a score or rank for this model on the hidden test set.

not tested on the hidden test set[1]. Detailed scores for local 5-fold cross validation and scores on hidden validation set are shown in Table 2.

## 4. Discussion and Conclusions

In the prior PhysioNet competition we attained a 5-fold cross-validation metric score of 0.684 on 12-lead ECG data alone [3]. However, the current model has greater potential for application and generalisability thanks to its ability to extract domain-invariant features from 12-lead or reduced-lead ECGs.

Although the domain-invariant model showed improvement on the unseen domain when compared to the baseline model, both models had 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. In future work, we will apply this model to the test set, and systematically optimise hyperparameters.

## References


[1] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[2] Alday EAP, Gu A, Shah AJ, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/computing in cardiology challenge 2020. Physiological Measurement jan 2021;41(12):124003.

[3] Zhao Z, Fang H, Relton SD, Yan R, Liu Y, Li Z, et al. Adaptive Lead Weighted ResNet Trained with Different Duration Signals for Classifying 12-lead ECGs. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.


[1] the baseline model was tested, and received scores of 0.42 (12 lead), 0.42 (6 lead), 0.42 (4 lead), 0.44 (3 lead) and 0.38 (2 lead) the average over all leads was 0.41


[4] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics 2018;8(7):1368—1373.

[5] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. PhysioBank PhysioToolkit and PhysioNet 2008;Doi: 10.13026/C2V88N.

[6] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik 1995;40(S1):317–318.

[7] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[8] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. Scientific Data 2020;7(48):1–8.

[9] Zheng J, Cui H, Struppa D, Zhang J, Yacoub SM, El-Askary H, et al. Optimal Multi-Stage Arrhythmia Classification Approach. Scientific Data 2020;10(2898):1–17.

[10] 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.

[11] Engelse WA, Zeelenberg C. A Single Scan Algorithm for QRS-detection and Feature Extraction. Computers in Cardiology 1979;6(1979):37–42.

[12] 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.

[13] 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.

[14] 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.

[15] 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