Multi-Label Cardiac Abnormality Classification from Electrocardiogram Using Deep Convolutional Neural Networks

Nima L Wickramasinghe[1], Mohamed Athif[2]

[1]Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka
[2]Department of Biomedical Engineering, Boston University, Boston, USA

Abstract

This paper proposes a deep neural network architecture to perform multi-label classification of 26 cardiac abnormalities from 12-lead and reduced lead ECG data. The model was created by team "NIMA" for the PhysioNet/Computing in Cardiology Challenge 2021. ECG signals of at most 20 seconds in length were used for training. The data are preprocessed by normalizing, resampling, and zero-padding to get a constant-sized array. The preprocessed ECG signals and Fast Fourier Transforms obtained from the preprocessed signals are each fed into two separate deep Convolutional Neural Networks. Spatial dropouts and average pooling are used between each convolutional layer to reduce overfitting and to reduce model complexity. Following the convolutional layers, the time and frequency domain network outputs are concatenated and passed through two dense layers that output an array of size 26. A threshold of 0.13 is used on the output array to determine the class while addressing data imbalance. The method achieved a score of 0.55, 0.51, 0.56, 0.55, and 0.56 ranking 2nd, 5th, 3rd, 3rd and 3rd out of 39 officially ranked teams on 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead hidden test datasets, respectively, according to the challenge evaluation metric. Our model performs comparable to the 12 Lead ECG using smaller subsets of leads.

1. Introduction

Cardiovascular diseases have a high prevalence (in the USA for 49.5% of adults >20 years) and are a major cause of death [1]. An electrocardiogram (ECG) measures the heart's electrical activity and helps the diagnosis of cardiovascular diseases by identifying various abnormalities in patterns. Automatic detection of cardiac abnormalities from ECGs has many benefits, including early detection and better prognosis. While such implementations often use the standard 12-lead ECGs, using a smaller number of leads would enable low-cost, portable, and user-friendly point of care devices. However, it remains largely unexplored whether similar outcomes are achievable using reduced leads. The objective of The PhysioNet/Computing in Cardiology Challenge 2021 was to find automated, open-source approaches to identify multiple cardiovascular diseases from 12-lead and reduced-lead ECG data [2, 3].

Deep learning methods have recently gained popularity in classifying various cardiac abnormalities in addition to traditional methods such as support vector machines, linear regression, decision trees, and feed-forward neural networks [4, 5]. In this work, we propose and investigate the suitability of a deep learning-based method using the time domain and frequency domains of ECG signals to classify 26 different classes of cardiac abnormalities using 12-lead and reduced-lead ECG data.

2. Methods

2.1. Dataset

Local training is done using the CPSC [6], PTB [7], PTB-XL [8], INCART [9], Chapman-Shaoxing [10], Ningbo [11] and Georgia databases with 88,259 12-Lead ECG signals. Recordings with a length greater than 20 seconds are not used for training (<3% of the dataset). The final dataset containing 85,811 records is randomly divided into a training dataset (79,791 records) and a local validation dataset (6,020 records). Reduced lead ECG signals are generated by selecting the required leads from 12-lead ECG signals as shown in Table 1.

Leads	Leads used
12	I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6
6	I, II, III, aVR, aVL, aVF
4	I, II, III, V2
3	I, II, V2
2	I, II

Table 1. Leads used in different lead sets

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.352

2.2. Preprocessing

Firstly, all the recordings are normalized by subtracting the baseline and dividing by the Analog-to-Digital Converter (ADC) gain. Recordings that are less than 20 seconds are zero-padded to 20 seconds and then resampled at 200Hz, giving an array with 4,000 time points. Fast Fourier Transform (FFT) is applied to the resulting recordings giving a complex array of 4,000 points representing the frequency domain of the recording. Since the frequency domain of a real signal is symmetric, only the first 2,000 points are selected. The magnitude and phase of the resulting array are separated and concatenated, giving an array of shape (2N, 2,000), where N is the number of leads. The time-domain array is again resampled at 100Hz, giving an array of shape (N, 2,000). The 26 classes of diagnoses are one-hot-encoded to give a binary array of size 26.

SciPy library is used for resampling and obtaining the FFT of the signals. Both the time domain and the frequency domain arrays are sent as inputs to the deep learning model.

2.3. Model Description

The model is created using the Python TensorFlow library using the Keras functional API.

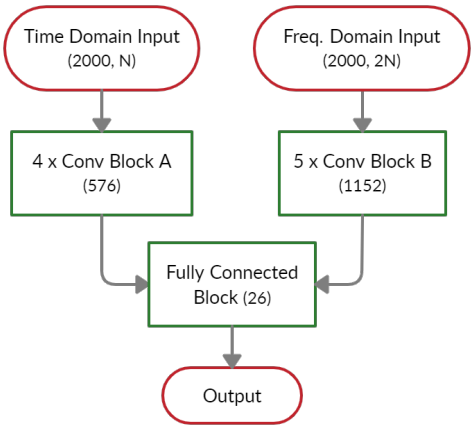

Figure 1. The neural network architecture. The numbers in brackets show the shape of the output array from the particular block. N = number of leads

The time-domain input is sent into convolution block A (Figure 2). It is first passed through a 1D convolutional layer, and then Swish activation function [12] is applied to the outputs. A skip connection is used to pass the information from the previous layer directly to the next layer. Values from both the skip connection and the convolutional layer are added, and Swish activation function is applied to the outputs. Then we used 1D spatial dropout [13] with a rate of 0.2 to reduce the overfitting of the model. The outputs are passed through an average pooling layer to reduce the dimensions of the feature maps and reduce computational cost. This process is repeated four times for the time domain signal with the number of filters increasing by two-fold and the kernel size increasing by 2 each time as shown in Table 2. Finally, instead of the average pooling layer, a global average pooling layer is used to produce an output array of size 576. We used multiples of the number of leads for the number of filters used in the convolutional layers, a large initial kernel size, and padding in all the convolutional layers to maintain consistent dimensionality across convolutional layers and to facilitate skip connections.

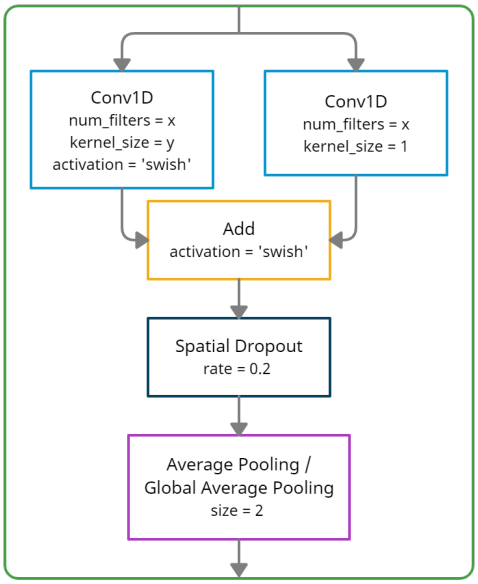

Figure 2. Convolution block A

The frequency-domain input is sent into convolution block B (Figure 3). It is first passed through a 1D convolutional layer, and then the Swish activation function is applied to the resulting output. Spatial dropout is then applied with a rate of 0.1 and then average pooled with a size of 2. This is repeated 5 times with the number of filters increasing by 2 fold and the kernel size increasing by 2 with each repetition as shown in Table 2. Finally, a global average pooling layer is used, producing an output array of size 1,152.

The outputs from convolution block A and convolution block B are concatenated (Figure 3), and Swish activation function is applied. The resultant output is fed into a fully connected layer with 576 units. A dropout layer with a rate of 0.5 is applied, and another dense layer of 26 units with

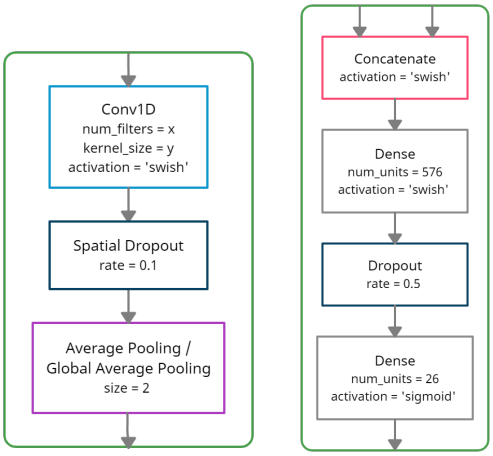

Figure 3. Convolution block B (Left) and fully connected block (Right)

sigmoid activation function finally outputs an array of size 26, which is used to predict whether the specific cardiac abnormality is present or not. Our method classifies the 26 scored classes, including the 4 pairs of similar diagnoses as 4 classes and is able to identify multiple diagnoses from a single recording.

The loss and Area Under the Precision-Recall Curve (AUPRC) are used as metrics when evaluating the model performance while training. Training is discontinued if the validation AUPRC values did not increase by 0.01 for 5 epochs. The data is shuffled between each epoch and a batch size of 128 is used for training. First, the time domain and frequency domain network hyperparameters were optimized independently and then fine-tuned together. We used a grid search method to find the optimum hyperparameters. Different lead components are used as the number of channels when used as input to the convolutional layers. Only the number of input channels are changed in the model when different subsets of leads were used.

Block	A1	A2	A3	A4	B1	B2	B3	B4	B5
Num. of filters (x)	72	144	288	576	72	144	288	576	1152
Kernel size (y)	15	3	5	7	3	5	7	9	11

Table 2. Hyperparameters used in the final model. A and B stand for the convolutional block A (Figure 2) and B (Figure 3) respectively. The number following A or B indicates the repetition of each block.

Binary cross entropy is used to measure the loss. We used Adam optimization with an initial learning rate of 0.001 for all the models. The learning rate is reduced tenfold after a number of epochs as shown in Table 3.

Lead set	12-lead	6-lead	4-lead	3-lead	2-lead
Epoch at which LR=0.0001	17	23	16	16	15
Total number of epochs	21	26	22	18	17

Table 3. Details containing the number of epochs each lead set is trained

2.4. Model Evaluation

When evaluating, if the length of the recording is larger than 20 seconds we segmented the recording into 20-second windows sequentially with no overlap. If the last segmented window is less than 5 seconds, it is discarded; otherwise, it is zero-padded to 20 seconds. All the segmented windows are sent as input to the trained model. If the cardiac abnormality is found in at least one of the windows, then the cardiac abnormality is marked as present. We empirically determined the optimal threshold to be 0.13 by changing the threshold between 0 and 1 with a step of 0.01 and validating against the training data using the challenge scoring metric.

3. Results

Table 4 shows the challenge score obtained by the above models trained on different leads on the local validation set, hidden validation set and, the hidden test set along with the ranking.

Leads	Training	Validation	Test	Ranking
12	0.75	0.65	0.55	2
6	0.70	0.59	0.51	5
4	0.72	0.63	0.56	3
3	0.73	0.63	0.55	3
2	0.70	0.61	0.56	3

Table 4. Challenge scores for the final accepted entry (team NIMA) on the local validation set, hidden validation set and, the hidden test set along with the ranking

Table 5 shows the F1 scores obtained by each class on the local validation set.

4. Discussion and Conclusions

The classifier performs with scores ranging from 0.59 - 0.65 in the hidden validation set. Overall, the 6-lead and 2-lead sets that did not contain the V2 lead show the lowest score. Therefore, V2 lead seems to play an important role in classifying cardiac abnormalities. The performance on the hidden test set is lower than the hidden validation set but similar across all lead sets.

Multi-label data imbalance is especially harder to address because upsampling one of the rare classes may re-

Dx	F1 Score	Dx	F1 Score
AF	0.59	PAC/SVPB	0.57
AFL	0.83	PR/VPB	0.89
BBB	0.28	PRWP	0.31
Brady	0.62	PVC	0.63
CLBBB/LBBB	0.75	LPR	0.46
CRBBB/RBBB	0.85	LQT	0.48
IAVB	0.68	QAb	0.35
IRBBB	0.51	RAD	0.62
LAD	0.72	SA	0.68
LAnFB	0.72	SB	0.96
LQRSV	0.50	STach	0.94
NSIVCB	0.38	TAb	0.59
NSR	0.91	TInv	0.49

Table 5. F1 scores obtained by the final model for 26 scored diagnoses on the local validation set

sult in oversampling one of the abundant classes and vice versa. Upsampling and downsampling data both as individual classes and as super sets of all 26 classes to create balanced datasets did not yield better performance. Therefore, to address the class imbalance, we used a single optimized threshold on the sigmoid output.

Having a shorter input array size reduced the training time of the model significantly while preserving informative features in the signals. Therefore, we resampled the 20 second long recordings at 100Hz. We did not include any filtering steps to clean the ECG since it may eliminate discriminating features between classes. Further investigation is required on whether filtering improves performance.

Model variation	Score
Final Model	0.745
Replacing Spatial dropout with Dropout	0.733
Without using the frequency domain	0.734
Without using the time domain	0.679

Table 6. Challenge scores for different variations of our final model on the local validation set

Using the frequency domain as input additional to the time domain input, using spatial dropouts, using a number of convolution layer filters in multiples of the number of leads, using large initial kernel sizes, using Swish activation instead of Rectified Linear Unit activation, and using residual networks through skip connections improved the performance and stability of the network (Table 6).

Acknowledgments

The authors would like to thank the Sustainable Education Foundation, Sri Lanka for facilitating the collaboration and Richie Wheelock for proofreading.

References

[1] Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics – 2021 Update: a Report from the American Heart Association. Circulation 2021;143(8):e254–e743.

[2] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2020;41.

[3] 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.

[4] Aurore L, Ana M, Pablo MJ, Pablo L, Blanca R. Computational Techniques for ECG Analysis and Interpretation in Light of their Contribution to Medical Advances, 2018.

[5] Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic Diagnosis of the 12-lead ECG using a Deep Neural Network, 2020.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Ramachandran P, Zoph B, Le QV. Searching for Activation Functions, 2017.

[13] Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient Object Localization Using Convolutional Networks, 2015.

Address for correspondence:

Nima L. Wickramasinghe
195/2, Galgediyawa, Gampola, Sri Lanka
wickramasinghenlssck@gmail.com