# Channel Self-Attention Deep Learning Framework for Multi-Cardiac Abnormality Diagnosis from Varied-Lead ECG Signals

Apoorva Srivastava[1], Ajith Hari[1], Sawon Pratiher[1], Sazedul Alam[2], Nirmalya Ghosh[1], Nilanjan Banerjee[2], Amit Patra[1]

[1] *Department of Electrical Engineering, Indian Institute of Technology Kharagpur, West Bengal, India.*
[2] *Computer Science and Electrical Engineering, University of Maryland-Baltimore County, Maryland, USA.*

**Abstract**

*Electrocardiogram (ECG) signals are widely used to diagnose heart health. Experts can detect multiple cardiac abnormalities using the ECG signal. In a clinical setting, 12-lead ECG is mainly used. But using fewer leads can make the ECG more pervasive as it can be integrated with wearable devices. At the same time, we need to build systems that can diagnose cardiac abnormalities automatically. This work develops a channel self-attention-based deep neural network to diagnose cardiac abnormality using a different number of ECG lead combinations. Our approach takes care of the temporal and spatial interdependence of multi-lead ECG signals. Our team participates under the name* ***"cardiochallenger"*** *in the "PhysioNet/-Computing in Cardiology Challenge 2021". Our method achieves the challenge metric score of 0.55, 0.51, 0.53, 0.51, and 0.53 (ranked $2^{nd}, 5^{th}, 4^{th}, 5^{th}$ and $4^{th}$) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead cases, respectively, on the test data set.*

## 1. Introduction

With over 17.9 million deaths, cardiovascular diseases are the leading cause of mortality worldwide [1]. The heart's activity from different angles can be studied from a 12-lead ECG. Detection of multiple cardiac abnormalities like coronary occlusion, myocardial infarction, etc., can be done using a 12-lead ECG.

Early-stage prognosis and timely interventions aid clinicians in identifying different cardiac irregularities and provide improved clinical outcomes. The PhysioNet/CinC-2021 challenge is dedicated to cardiac abnormality classification (CAC) from 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECG recordings [2]. Early and accurate detection of diseases with fewer leads makes ECG greater pervasive as it can be incorporated with wearable devices. Conventional CAC methods regularly employ machine learning models on the extracted domain-aware handcrafted features using raw ECG signal processing. Of late deep learning (DL) methods have democratized the CAC task with superior performance [3],[4],[5]. DL models can abstract explanatory ECG feature representations in an automated fashion and predict CACs in an end-to-end manner [6],[7]. This paper proposes an attention-based DL model, which will help medical practitioners judiciously inspect and categorize the inter-beat and intra-beat patterns. The proposed model acknowledges the spatial interrelation among the channels and the important temporal segments of the signal.

The rest of the paper is organized as follows. Section 2 summarizes the data pre-possessing and our channel self-attention-based DL model. Experimental results are discussed in sections 3 and 4. Section 5 concludes the paper.

## 2. Methodology

Cardiac abnormality detection using ECG signals can be formulated as a time-series classification problem. We aim to detect 29 multi-labeled cardiac abnormalities along with sinus rhythm using varying lead ECG signals [2]. The model is trained on 12-lead ECG and tested on:

- 12-lead: I,II,III,aVR,aVL,aVF,V1,V2,V3,V4,V5,V6
- 6-lead: I,II,III,aVR,aVL,aVF
- 4-lead: I,II,III,V2
- 3-lead: I,II,V2
- 2-lead: I,II

In this paper, a channel self-attention (CA)-based framework, as depicted in Figure 1 is proposed for the diagnosis of multi-labeled cardiac abnormalities. The model is inspired by squeeze and excitation network [8]. The global spatial information is squeezed, and channel-wise statistic is generated by CA framework. Higher weight is given to the more imperative channel, which leads to enhanced performance. Here, it is applied with the inception and residual neural model. In the following section, we provide a detailed description of the system's components.

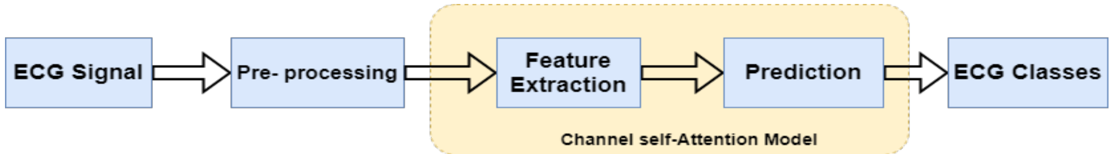


Figure 1. Pipeline for cardiac abnormalities detection

## 2.1. Data pre-processing

The publicly available challenge dataset consists of 88,253 twelve-lead ECGs recordings. The data is collected from 4 countries across 3 continents. The sampling frequency of the ECG signals varies from 257 Hz to 1000 Hz, and signal duration ranges from 6 seconds to 30 minutes. All the signals are down-sampled to 125 Hz to handle the variable sampling frequency of the signals. If the signal is too long, it is truncated after 120 seconds.

## 2.2. Channel self-attention based DL model

The input to the proposed CA-based model is a variable-length ECG segment $(X = [x_1, ...., x_k])$ and the prediction of 29 cardiac abnormalities along with sinus rhythm is the model's output. The proposed CA-based architecture is shown in figure 2(a). The channel depth and the number of inception and residual blocks are experimentally decided, and the tuning of network parameters is done by hit and trial. The channel self-attention-based DL model is the ensemble of: 1) Inception and Residual architecture; 2) Channel self-Attention architecture; 3) Attention pooling.

### 2.2.1. Inception and residual architecture

The idea of inception came from [9], where sparsely connected architecture was introduced to replace the fully connected connection of Convolution Neural Network (CNN) layers. In this paper, the Inception model has convolution layers with a 1-D filter of sizes 3,4 and 5 with ReLu activation. The inputs to the CNN can be of a variable length, so the model is accustomed to handle variable-length data. The channel number increases as we move forward in the architecture. Figure 2(c) demonstrated the interior of the Inception blocks used in the model. The residual neural network helps in solving the problem of vanishing gradient [10]. The CNN layers in the Inception block maps the input $x_j$ to low dimension embedding $h_k = f_\psi(x_j)$, where $f_\psi$ is transformation function with parameter $\psi$. The output of the residual block is y = $\mathcal{F}$ $(h_j)$ +$\mathcal{G}(x_j)$, where $\mathcal{F}(.)$ shows the residual mapping to be learned and $\mathcal{G}(.)$ is the convolution layer added to match the dimension. Figure 2(b) demonstrates the residual block in the proposed DL model.

### 2.2.2. Channel self-attention architecture

CNN extracts the spatial features. Each channel represents the information of the feature map extracted. Adaptive weights can be assigned to the channel to find the interrelation among the channels. Therefore, we built a channel self-attention module to use the interdependence among the channels.

The idea of Channel self-Attention is derived from SE-Net [8] where inter-dependency among channels is captured as a function of channel description (global average). The main difference is that instead of finding explicit relation among channel descriptions, multiple channel descriptions are used, and channel attention is computed only using the feature vector extracted from the corresponding channel. The feature vector is extracted by passing spatial features channel-wise to one-dimensional CNN. 32 filters are used to extract 32 deep features as shown in figure 3. Here the interrelation among channels is captured through sharing the weights for calculating self-attention. Application of attention mechanism channel-wise can be regarded as the method of choosing semantic attributes.

### 2.2.3. Attention pooling layer

The attention pooling was introduced by [11] which is an adaptive multi-instance pooling method. Depending on the number of classes, it is modified to multi-head attention. Corresponding to every feature vector, segment weights are generated by multi-head attention neural network. The Softmax activation function is used to ensure that the weight sums up to 1. The attention mechanism helps make the model interpretable and replaces the widely used Long Short-Term Memory (LSTM) or Bi-LSTMs. The higher weights give importance to that segment of the signal.

Let $V = \{v_1, v_2, \cdots, v_K\}$ is a bag of $K$ feature vectors, then attention pooling is defined as: $p = \sum_{k=1}^{K} a_k v_k$. Here, $p \in \mathbb{R}^{\mathbb{N}\times\mathbb{L}}$ is the feature vector corresponding to $N$ classes heart disease. $a_k = \frac{exp\{W^T\ tanh(Uv_k^T)\}}{\sum_{j=1}^{K} exp\{W^T\ tanh(Uv_j^T)\}}$, where, $U$ and $W$ are trainable parameters.

## 2.3. Threshold optimisation

The output from the attention pooling layer is passed to the prediction layer for the detection of abnormalities.

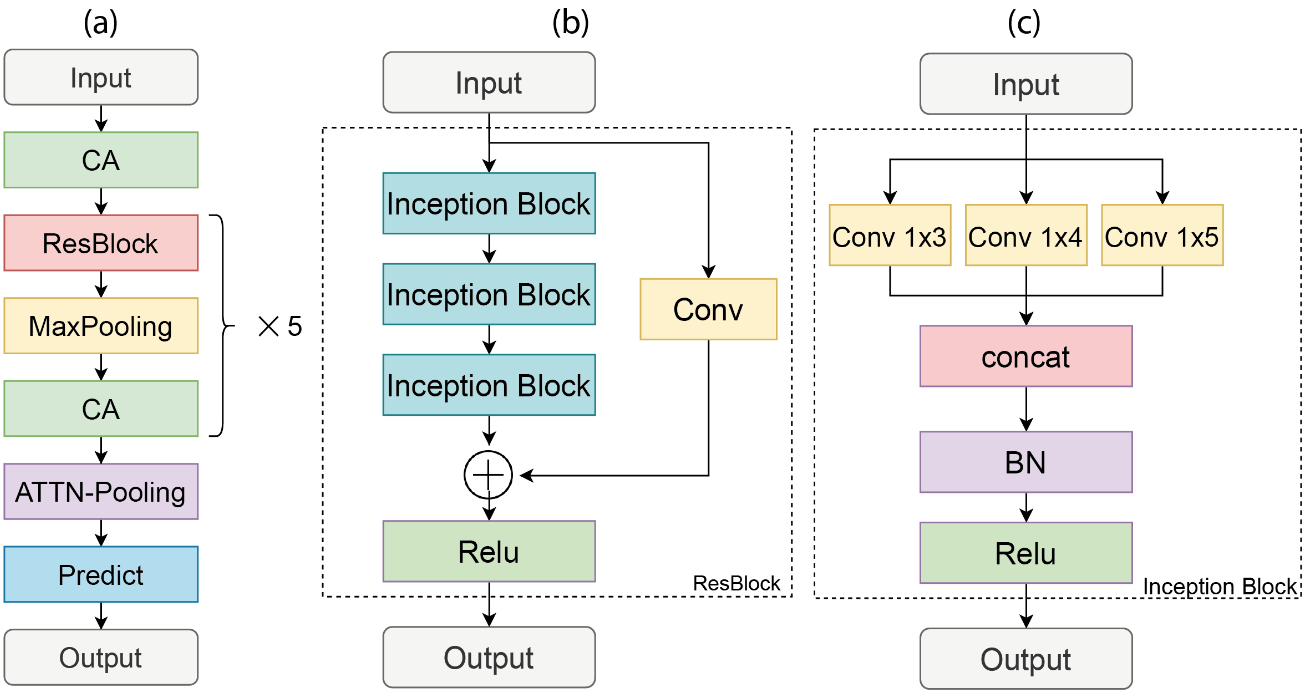


Figure 2. Channel self-Attention-based DL architecture. One-hot encoding of cardiac abnormalities classes is the output (grey) of the model. The figure illustrates: (a) proposed CA-based DL model; (b) Residual network block; (c) Inception network block.

The Sigmoid activation in the prediction layer will give the probabilities of the occurrence of sinus rhythm and 29 cardiac abnormalities. For evaluating the challenge metric score, these probabilities need to be changed in binary format by applying the threshold value on these predictions. If the prediction crosses the threshold value, 1 is assigned to the corresponding class else 0. The genetic algorithm optimized thresholds for each class that maximizes the challenge metric on the validation dataset.

## 2.4. Implementation Details

The dataset consists of 133 abnormalities, but in the challenge, it is required to detect 29 cardiac abnormalities along with sinus rhythm, and their SNOMED CT codes are included in the challenge evaluation metric along with the reward matrix ($W_{reward}$) [12]. The loss function is described below:

$loss(x, y) = mf \cdot BCE(x, y) - S_{normalised}$. Here, $BCE$ is Binary Cross-Entropy loss, $mf$ is a multiplication factor which scales the BCE loss by a factor of 0.1 if the difference between true and predicted label is less than 0.3 and $S_{normalised}$ is the normalised challenge metric which is computed as, $S_{normalised} = \frac{S_{observed} - S_{inactive}}{S_{true} - S_{inactive}}$, where, $S_{(x,y)} = X^T \times (\frac{y}{norm}) \cdot W_{reward}, norm = max(x + y - xy, 1)$.

For training, the model Adam optimizer was used with a learning rate of 0.001. The parameters are initialized using Xaviar uniform initializer. The early stopping method is also incorporated in the algorithm to avoid overfitting the model. The model trained for 100 epochs with 32 batch size. The model has 7,15,512 trainable parameters.

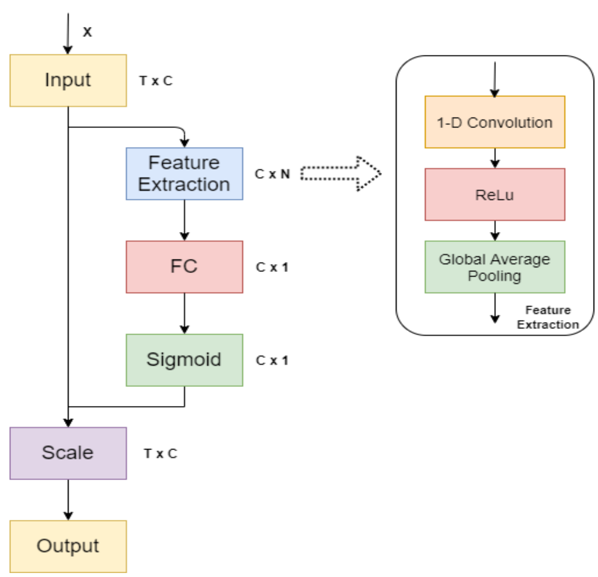


Figure 3. Schema of Channel self-attention architecture

## 3. Results

The table 1 shows challenge metric score on the train, validation and test dataset for variable number of ECG leads. For the training set, the proposed model is trained on a publicly available dataset given by PhysioNet/CinC challenge is trained and validated on 80% and 20% of training set respectively. The model is also validated and tested on hidden 6,630 and 16,630 ECG recordings. The proposed model requires 2497 minutes for training.

Table 1. Challenge metric score on training, validation and test set and overall ranking on hidden test set

| Leads | Training set | Validation set | Test Set | Ranking |
|---|---|---|---|---|
| 12-Lead | 0.81 | 0.64 | 0.55 | $2^{nd}$ |
| 6-Lead | 0.79 | 0.64 | 0.51 | $5^{th}$ |
| 4-Lead | 0.77 | 0.64 | 0.53 | $4^{th}$ |
| 3-Lead | 0.77 | 0.63 | 0.51 | $5^{th}$ |
| 2-Lead | 0.75 | 0.63 | 0.53 | $4^{th}$ |

## 4. Discussion

This paper proposed a channel self-attention DL model to detect cardiac abnormalities using 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead. The attention mechanism was used to capture the informative segment of the signal along with spatial interdependence among the channels.

We truncated the ECG data that is more than two minutes long during the pre-processing of the raw signal. Our decision to trim affected 74 samples, which is relatively low ($<$ 0.01 % of the total samples).

The duration of the input signal varies from 6 seconds to 2 minutes. We made the input length to our CA model a variable length to make it robust for any length of data to be used in real-time. We tried to make a generalized model that can handle the training dataset's heterogeneity and variable-length data.

## 5. Conclusion

This paper has described a channel self-Attention neural network-based approach to classify cardiac abnormalities presented in the PhysioNet/Computing in Cardiology Challenge 2021. Our DL model can classify 29 cardiac abnormalities and sinus rhythm with a challenge metric score of 0.55, 0.51, 0.53, 0.51, and 0.53 for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead cases, respectively, placing us in the top-five team in the competition. In future, we aim to handle data imbalance and infuse ensemble framework to enhance the model performance.

## References


[1] https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 (Accessed on 03/01/2021).

[2] M. A. Reyna, N. Sadr, E. A. P. Alday, A. Gu, A. J. Shah, C. Robichaux, A. B. Rad, A. Elola, S. Seyedi, S. Ansari, H. Ghanbari, Q. Li, A. Sharma, and G. D. Clifford, "Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021," *Computing in Cardiology*, vol. 48, pp. 1–4, 2021.

[3] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms using a Deep Neural Network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[4] A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr *et al.*, "Automatic Diagnosis of the 12-lead ECG using a Deep Neural Network," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020.

[5] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: A Systematic Review," *Computers in Biology and Medicine*, p. 103801, 2020.

[6] A. Natarajan, Y. Chang, S. Mariani, A. Rahman, G. Boverman, S. Vij, and J. Rubin, "A Wide and Deep Transformer Neural Network for 12-lead ECG Classification," *Computing in Cardiology*, pp. 1–4, 2020.

[7] A. Kirodiwal, A. Srivastava, A. Dash, A. Saha, G. V. Penaganti, S. Pratiher, S. Alam, A. Patra, N. Ghosh, and N. Banerjee, "A Bio-toolkit for Multi-Cardiac Abnormality Diagnosis Using ECG Signal and Deep Learning," *Computing in Cardiology*, pp. 1–4, 2020.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[11] M. Ilse, J. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," *International Conference on Machine Learning*, pp. 2127–2136, 2018.

[12] E. A. P. Alday, A. Gu, A. J. Shah, C. Robichaux, A.-K. I. Wong, C. Liu, F. Liu, A. B. Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G. D. Clifford, and M. A. Reyna, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiological Measurement*, vol. 41, no. 12, p. 124003, 2020.



Address for correspondence:

Apoorva Srivastava
EE, IIT Kharagpur, Kharagpur, West Bengal 721302, India
apoorva.srivastava.23@iitkgp.ac.in