Cardiac Abnormalities Recognition in ECG Using a Convolutional Network with Attention and Input with an Adaptable Number of Leads

Tomas Vicar[1], Petra Novotna[1], Jakub Hejc[2], Oto Janousek [1], Marina Ronzhina[1]

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communications, Brno University of Technology, Brno, Czech Republic
[2] International Clinical Research Center, St. Anne's University Hospital, Brno, Czech Republic

Abstract

In this work, we present an algorithm for automatically identifying the cardiac abnormalities in ECG records with the various number of leads. The algorithm is based on the modified ResNet convolutional neural network with the attention layer. The network input is modified to allow using a single network for different lead subsets. In an official phase challenge entry, our BUTTeam reached the 25-30th place. In our challenge entry, we have achieved 0.514, 0.500, 0.507, 0.508, and 0.498 of the challenge metric for 12, 6, 4, 3 and 2 leads, respectively. For evaluation on the training dataset (training-split), this metric reached 0.648, 0.633, 0.641, 0.641, and 0.634. From additional evaluation of other lead subsets, the leads representing a common heart axis orientation achieved the best detection results. However, all lead subsets (even a single lead method) performed very similarly.

1. Introduction

Deep learning-based ECG classifiers speed up diagnostics, save medical specialists' time and reduce human diagnostic inconsistency. The recent boom of low-cost and easy-to-use reduced-lead ECG systems challenges us to develop AI-based ECG classification working with fewer leads but achieving nearly the same results as standard 12-lead ECG classifiers.

Here we present the results of ECG classification into 27 categories using different combinations of ECG leads. These 1- to 12-lead combinations were made based on different principles. The first set of combinations was straightforwardly made according to the lead system types (e.g., selected chest and limb leads and combinations of both). For other combinations, we selected the leads with presumed benefits in the representation of specific anatomy-physiological phenomena (e.g., lead II is able to capture the P wave of variable morphology and chest leads are known to be helpful in ventricular arrhythmia diagnostics [1]). Finally, 1-lead systems could be one of the most interesting in wearable applications due to low computational demands, whereas multiple-lead ones could be more accurate in case of multi-class problems on one side and could be highly redundant on the other side. Compared to other studies in this field (e.g., [2–4]), our solution allows simple and fast training of the models due to the adaptable input attention layer. The models were verified on an extensive number of different ECG records.

2. Material and Methods

2.1. Data and preprocessing

This paper is a part of PhysioNet/Computing in Cardiology Challenge 2021 – "Will Two Do? Varying Dimensions in Electrocardiography" [5–7]. The training dataset consists of 12-lead ECG signals with 27 diagnoses to classify. For evaluation, datasets with the synthetically reduced number of leads are created (with 2/3/4/6/12-leads as shown in the first column of Table 1). Total 88,253 ECGs were shared publicly as training data, 6,630 ECGs were retained privately as validation data, and 16,630 ECGs were retained privately as test data.

The data prepossessing consisted of resampling, 50/60 Hz filtration, baseline wandering filtration and replication of the signals to achieve a fixed length. Resampling to 150 Hz was performed with linear interpolation and anti-aliasing FIR filtration. Undesired 50 and 60 Hz were removed via a second-order IIR notch filter [8] and baseline wandering was eliminated by a subtraction of a moving average weighted with 6 s long Blackman window [9]. Too short signals were replicated to achieve the 15 s length, whereas too long signals (4.2 %) were excluded from the training set. However, during the test phase, the long ECGs are split into 15 s long segments. The segments are analysed by the model separately and then the results are merged in such a way that all pathologies found in any segment of ECG are concerned in the final prediction.

2.2. Model Architecture

The proposed 1D Convolutional Neural Network (CNN) architecture is based on a residual neural network (ResNet) [10]. The model architecture is shown in Figure1. From the manually tested settings of the hyper-parameters, the following have shown to be the most efficient from the classification performance and computational time requirements (according to the challenge rule) point of view: the number of filters in the first layer $n = 24$, number for residual convolutional sub-blocks $L = 3$, and number o blocks $K = 7$. Moreover, we have included multiple heads for the adaptive number of leads and attention layers on the network output, which are described in the following section.

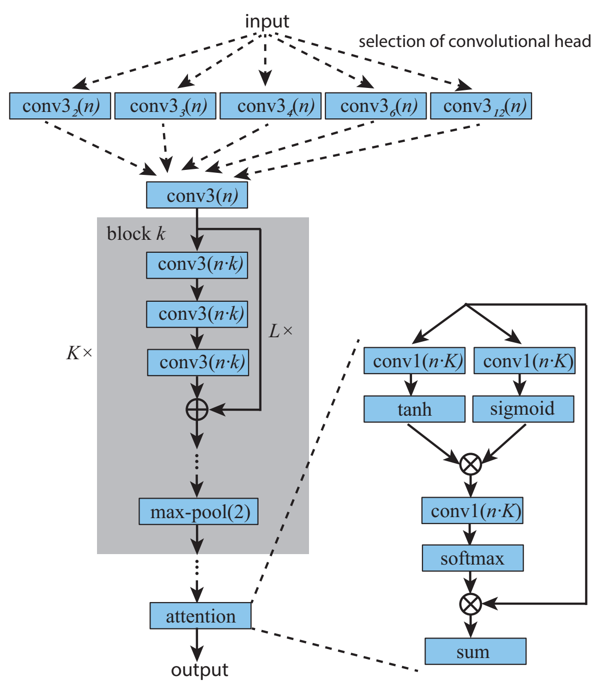

Figure 1: The architecture of proposed model with an adaptive number of leads on the input and attention on the output. k – block number of ResNet; K – number of blocks; L – number of convolutional sub-blocks with residual connection; n – number of filters in first layer; conv3(n) is convolution of filter size 3 and n filters. Each conv3 stands for a layer with convolution, batch normalization and ReLU non-linearity.

2.3. Adaptive number of leads and output attention

In the standard setting, different networks have to be trained for different number of ECG leads. Such networks, however, differ from each other only in the number of convolutional weights in the first layer, whereas the rest of the layers are the same. This opens the opportunity to "share" the weights between lead-specific models. One possible way is to share the whole network, but to replace the first layer and retrain the model for a particular number of leads. This will result in faster training due to initialisation from the pretrained network [11]. However, this solution still results in producing the number of models equal to the number of lead combinations. Our original solution allows to train a single network for the various number of leads simultaneously. This is shown in Figure1, where all convolution layers (convolutional heads) for all lead configurations are kept within the network and for each signal, a suitable head is applied. For each training iteration, a batch of signals with various randomly generated lead combinations is generated and used to train the model.

In the last layer of our network, an attention mechanism [12] was adopted in order to aggregate information from the whole signal. In addition to the classification improvement, the attention mechanism also generates so-called attention map, which can be used for the network decision interpretation. Compared to the original definition from [12]), in our case, the fully connected layers for all positions can be calculated effectively with the convolution of filter size 1, which leads to the calculation shown in Figure1.

2.4. Losses and implementation details

For the training of the proposed network, two different loss functions were used – weighted cross-entropy (WCE) and Challenge metric loss (CM) [13] specifically designed to minimise the differentiable version of the challenge metric from our previous work [13]. Training with CM loss seems to be unstable and leading to suboptimal results. Therefore, the network was first trained with WCE loss and then retrained with CM loss with learning rate for WCE 10^{-2}, 10^{-3}, 10^{-4} and for CM 10^{-2}, 10^{-3}, 10^{-4} and trained for 40, 20, 10 epochs and 20, 15, 10 epochs, respectively. The learning rates were manually adjusted for the optimal speed of convergence.

Adam optimizer with decoupled weight decay [14] ($\lambda = 10^{-5}$) was used for training with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and batch of size 64. Several data augmentation techniques were applied on the training set, including random circular shift, signal amplification along the voltage axis (up to $\pm$ 20 %) and signal stretch along the temporal axis (up to $\pm$ 10 %). The augmentation procedures were applied to the signals with the probability 0.8. The code is available at `https://github.com/tomasvicar/BUTTeam_ECG_classsification_CinC_2021`.

3. Results and Discussion

To evaluate the models within the design process, the available training dataset was divided into in-house train/validation/test sets in 8/1/1 ratio. Here, the evalua-

Table 1: Challenge metrics for different network settings and various combinations of ECG leads specified by challenge tasks. Training-split results correspond to the evaluation on a random subset of training data (in-house train/test/valid split). Validation results correspond to the output from the leaderboard of the official challenge phase. Test results were obtained on the final challenge test set.

	training-split			validation	test
leads	initial	multi-head	multi-head w/o att.	multi-head	multi-head (aval. after conf.)
I, II, III, aVR, aVL, aVF, V1-V6	.659	.648	.629	.514	.xxx
I, II, III, aVR, aVL, aVF	.638	.633	.614	.500	.xxx
I, II, III, V2	.644	.641	.625	.507	.xxx
I, II, V2	.641	.641	.623	.508	.xxx
I, II	.635	.634	.613	.498	.xxx

tion on the in-house test set, on the challenge validation set and on the challenge final test set is referred as training-split, validation and test, respectively. Results obtained for the initial network (i.e., several networks trained separately for each lead combination), multi-head network (i.e., one network trained simultaneously for different combinations of leads) and multi-head network without attention (i.e., the network with global max-pooling instead of attention layer) are summarised in Table 1. The multi-head network was trained with the same settings as the initial network including the same number of epochs. The only difference between the networks was in the batches, which consisted of signals with different lead sets. Thus, only a single network was trained instead of 5 separate networks, which took 5-times less time than the initial network training. Moreover, only a single, ”universal” network is required for evaluation. However, this leads to a small drop in network performance on training-split by 0.011, 0.005, 0.003, 0.000, and 0.000 for 12, 6, 4, 3 and 2 leads, respectively. The main drop in performance happened for 12 leads, whereas the performances of 3 and 2 lead settings were the same as for initial network. It was expected because for 2 leads (I and II), these leads are shared between various lead subsets resulting in sharing the extracted features. On the other hand, for 12 lead ECGs, specific features have to be extracted to use otherwise unused leads. Finally, we removed attention from our network to test its significance. It leads to a significant drop in network performance on training-split by 0.019, 0.019, 0.016, 0.018, and 0.021 for 12, 6, 4, 3, and 2 leads, respectively.

3.1. Leads analysis

Electrode placement may have a substantial impact on the amount of information we can retrieve under the condition of a limited number of ECG leads in scenarios such as home monitoring [15]. In order to exhaust the possibilities given by this year's challenge topic, we evaluated more than the required lead combinations to determine which one did the best work (see Figure2). Considering only single lead arrangements, the highest score of 0.622 was achieved by using lead II. This is in accordance with other studies [1,2], where a strong relationship between the main electrical axis (and, thus, electrophysiological properties of the heart) and lead II was shown. The 2nd and the 3rd best performing leads – aVR and aVF (with the score of 0.612 in both cases) – represent the electrical axis in similar way as the lead II. Puurtinen et al. in [16] stated that the area around the precordial leads V2, V3 and V4 and above V1 and V2 is the best for QRS complex and P wave detection. We, therefore, expected these leads to be helpful in identification of arrhythmias manifesting in these parts of the ECG. Interestingly, in our setup the best performing pair of precordial leads were those close to the anterior and left lateral side of the heart, with the highest score of 0.628 and 0.627 for pairs V4-V6 and V3-V5, respectively. Further improvement of the model performance was reached by adding ECG signals either from the remaining original Einthoven's or precordial leads. The best overall score of 0.670 was obtained for the combination of leads II, III, V1–V6. All those leads are known to be uncorrelated, which obviously helped the model to better learn ECG features. This hypothesis is further supported by the substantial worsening of the metrics by including the rest of the highly correlated leads aVR, aVL and aVF.

4. Conclusions

Our algorithm based on the modified ResNet CNN with the attention layer automatically identifies the cardiac abnormalities in 12, 6, 4, 3, and 2-lead ECG records. The input of the network is modified to allow using a single network for different lead subsets. Our BUTTeam reached 25-30 place and achieved 0.498-0.514 of the challenge metric in the official challenge phase entry and 0.633-0.648 for evaluation on the training dataset (training-split). The best results were obtained for the leads, which reflect the heart electrical axis orientation well. To conclude, any presented lead subset (even a single-lead setup) can be successfully used since they performed similarly; thus, any lead will do.

A

I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
.588	.622	.583	.612	.578	.612	.571	.574	.586	.609	.618	.610

B

V1, V2	V2, V3	V3, V4	V4, V5	V5, V6	V1, V3	V2, V4	V3, V5	V4, V6	V1, V4
.604	.608	.617	.627	.624	.613	.621	.627	.628	.624

C

I, II, III, aVR, aVL, aVF, V1 - V6	.659
I, II, III, aVR, aVL, aVF	.638
I, II, III, V2	.644
I, II, V2	.641
I, II	.635

D

I, II, III, aVR, aVL, aVF, V1 - V6	.659
I, II, III, aVR, aVL, aVF	.638
I, II, III	.643
I, II, III, V1-V6	.669
II, III	.640
II, III, V1-V6	.670

E

I, II, III, V1	.653
I, II, III, V3	.652
I, II, III, V4	.652
I, II, III, V5	.656
I, II, III, V6	.653

Figure 2: Tables of challenge metric results with various combination of leads. Results are calculated for initial network. Evaluation of A) single leads, B) pairs of precordial leads, C) challenge lead combination, D) without redundant leads, E) limb leads with single precordial lead.

References

[1] da S. Luz EJ, Schwartz WR, Cámara-Chávez G, Menotti D. Ecg-based heartbeat classification for arrhythmia detection: A survey. Computer Methods and Programs in Biomedicine 2016;127:144–164.

[2] Sujadevi V, Soman K. Towards identifying most important leads for ecg classification. a data driven approach employing deep learning. Procedia Computer Science 2020; 171:602–608. Third International Conference on Computing and Network Communications (CoCoNet'19).

[3] Yang W, Si Y, Wang D, Zhang G. A novel approach for multi-lead ecg classification using dl-ccanet and tl-ccanet. Sensors 2019;19(14).

[4] Llamedo M, Khawaja A, Martínez J. Analysis of 12-lead classification models for ecg classification. In 2010 Computing in Cardiology. 2010; 673–676.

[5] Alday EAP, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad AB, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the Physionet/Computing in Cardiology Challenge 2020. Physionet. Physiological Measurement 2020;(In Press).

[6] 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 2020; 41(12):124003.

[7] 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.

[8] Orfanidis SJ. Introduction to signal processing. Prentice-Hall, Inc., 1995.

[9] Blackman RB, Tukey JW. The measurement of power spectra from the point of view of communications engineering—part i. Bell System Technical Journal 1958; 37(1):185–282.

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

[11] Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on knowledge and data engineering 2009; 22(10):1345–1359.

[12] Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In International conference on machine learning. PMLR, 2018; 2127–2136.

[13] Vicar T, Hejc J, Novotna P, Ronzhina M, Janousek O. Ecg abnormalities recognition using convolutional network with global skip connections and custom loss function. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.

[14] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. arXiv preprint arXiv171105101 2017;.

[15] Zhu H, Pan Y, Wu F, Huan R. Optimized electrode locations for wearable single-lead ecg monitoring devices: A case study using wfees modules based on the lans method. Sensors 2019;19(20):4458.

[16] Puurtinen M, Viik J, Hyttinen J. Best electrode locations for a small bipolar ecg device: Signal strength analysis of clinical data. Annals of biomedical engineering 2009; 37(2):331–336.

Address for correspondence:

Tomas Vicar
Department of Biomedical Engineering,
Brno University of Technology,
Technicka 12, 616 00 Brno, Czech Republic.
vicar@vutbr.cz