Classification of ECG using Ensemble of Residual CNNs with Attention Mechanism

Petr Nejedly[1], Adam Ivora[1], Radovan Smisek[1], Ivo Viscor[1], Zuzana Koscova[1], Pavel Jurak[1], Filip Plesinger[1]

[1]Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

Abstract

Here in this paper, we introduce a solution to the PhysioNet Challenge 2021. The method is based on the ResNet deep neural network architecture with a multi-head attention mechanism for ECG classification into 26 independent groups. The model is optimized using a mixture of loss functions, i.e., binary cross-entropy, custom challenge score loss function, and sparsity loss function. Probability thresholds for each classification class are estimated using the evolutionary optimization method. The final model consists of three submodels forming a majority voting classification ensemble. The proposed method can classify ECGs with a variable number of leads, e.g., 12-lead, 6-lead, 3-lead, and 2-lead. The algorithm was trained and validated on the public dataset proposed for the challenge. The trained algorithm was tested using a hidden validation set during the official phase of the challenge and obtained validation scores (ISIBrno-AIMT): 0.64, 0.62, 0.63, 0.63, and 0.62 for lead configurations 12, 6, 3, 4, and 2, respectively. The total training time was approximately 27 hours, i.e., 9 hours per model.

1. Introduction

Cardiovascular diseases are the most common cause of death globally, reaching 32 percent by 2019 [1]. Heart disorders are usually analyzed using electrocardiographic signals (ECG) at a length of 10-60 seconds, acquired from the body surface. The ECG signal shows the electrical activity of heart atria and ventricles and, therefore, informs about heart rhythm and a beat morphology.

Current automated algorithms to analyze the ECG signal are based on machine-learning (using expert features) or deep-learning methods. A specialty of deep-learning methods is that they extract features by themselves during a training process from a raw or transformed ECG signal. These deep-learning methods are usually based on convolutional layers and are called convolutional neural networks (CNN). A need to train very complex CNNs led to the invention of the Residual Networks (ResNet) architecture [2], implementing residual blocks to improve gradient propagation training.

For this Computing in Cardiology 2021 challenge [3] we introduce a solution using an ensemble of a custom variant of ResNet neural networks accompanied by a multihead attention mechanism. This solution arises from our last year Computing in Cardiology 2020 challenge solution, where we were able to achieve an acceptable validation score, however, the final test scores showed poor performance while testing on the undisclosed testing database. This indicated that our method was able to classify data that originated from the same hospital very well, however, generalizability for other institutions was missing [4]. This year's solution tries to improve drawbacks from previous years by introducing several changes to our method.

The investigation of literature from previous year solutions [5] led us to change the preprocessing step by introducing data filtering to maximize generalizability across the institutions. The filtering should minimize the ECG frequency band as much as possible (for the cost of possibly discarding some ECG information that might be useful). We believe that this might be a good idea since we are not aware of data quality, types of artifacts, and distortions in undisclosed testing sets. Secondly, we utilize z-score normalization, while last year, we were using physical units in mV. In addition, models are trained using a custom loss function which consists of three parts, i.e., cross-entropy, custom challenge loss, and custom sparsity loss. The custom challenge loss optimization was proposed by [6], where the continuous equivalent of binary OR operator was used to design differentiable approximation of challenge metric. This helps the model to learn the similarity of diagnoses. Next, we introduce the custom sparsity loss, which forces the model to output probability values close to 0 or 1. This helps with the final threshold optimization to binarize the data output. Lastly, the class-specific thresholds are found using differential evolution genetic algorithms. The final model consists of three subunits creating the model ensemble.

2. Methods

For this challenge, we have introduced a fully autonomous cloud-based solution for training and deployment of deep-learning models utilizing publicly available Python libraries such as NumPy, SciPy, scikit-learn and PyTorch. For training and validation, the public challenge dataset was split into two sub-datasets in ratios 80 percent and 20 percent, respectively. The dataset stratification was iteratively optimized by a method available in scikit-multilearn based on [7].

The data preprocessing consists of several steps described below:

1. Provided data are expanded into fixed 12-lead configuration. If any lead is missing, the particular matrix row is filled with zeroes. This transformation always outputs a matrix with dimensions (12, time).
2. Resampling: Data are resampled to the sampling frequency of 500 Hz. Polyphase filtering is used when the original sampling frequency is 1000 Hz; otherwise, the FFT method is used for resampling.
3. Filtering: Data are filtered using a zero-phase method with 3rd order Butterworth bandpass filter with frequency band from 1 Hz to 47 Hz.
4. Normalization: Each ECG channel is normalized using a z-score.
5. Zeropadding: Data are zero-padded into the shape of 8192 samples in the time domain. If a signal length is larger than 8192, then the signal is randomly sampled and cut into the length of 8192.
6. Augmentation: During the training phase, randomly choose the lead configuration (e.g. 12, 6, 4, 3, 2). Leads that are not used are filled with zeros.

The model architecture is designed on the custom ResNet blocks that utilize large convolution sizes (1st conv layer 15 and subsequent residual conv layers 9 and using stride 2x). The output from the convolutional layers is forwarded through the multi-head attention mechanism and subsequently pooled with adaptive max pooling. The resulting feature vector is concatenated with a binary ECG lead indicator and classified by fully connected layers.

The model has two output heads. A first head outputs the logits forwarded into the BCE loss function. The second output forms an additional small neural network that processes logits from the first output head and optimizes challenge score and sparsity of probabilities, i.e., challenge loss and sparsity loss. The small network does not propagate gradients into the bottom layers. This means that the bottom layers of the model for feature extraction are optimized by BCE. And the top layer that outputs the probabilities is optimized by challenge and sparsity loss.

The model is trained using Adam optimizer for 50 epochs with learning rate 1e-3, batch size = 128, and L2 regularization parameter 1e-4 while reducing the learning rate by a factor of 0.1 after every 20th epoch. The optimization loss function is composed of three units i.e., binary cross-entropy (BCE), custom challenge loss (CL), and custom sparsity loss (SL). The custom challenge loss (differentiable approximation of challenge score) forces the network to maximize challenge score, accounting for class weights. The method was proposed by [6] during the previous year of challenge. In addition, we propose a sparsity loss derived from the parabolic curve that penalizes the network for outputting probability values that are close to 0.5 thus forcing it to output probability values close to 0 or 1, which helps with final threshold optimization.

$$Loss = \sum_{batch} BCE(t,p) - CL(t,p) + SL(p) \quad (1)$$

$$SL(p) = -4p(p-1) \quad (2)$$

$$CL(t,p) = \sum_{ij} w_{ij} a_{ij}(t,p) \quad (3)$$

In general, the inputs to the loss functions are targets t and probabilities p. The challenge loss is estimated from modified multi-class confusion matrix entries a_{ij} and its coresponding weights w_{ij}. The authors of [6] proposed to estimate normalization constant N for modified confusion matrix entries a_{ij} using continuous version of logic OR function, which makes loss function differentiable.

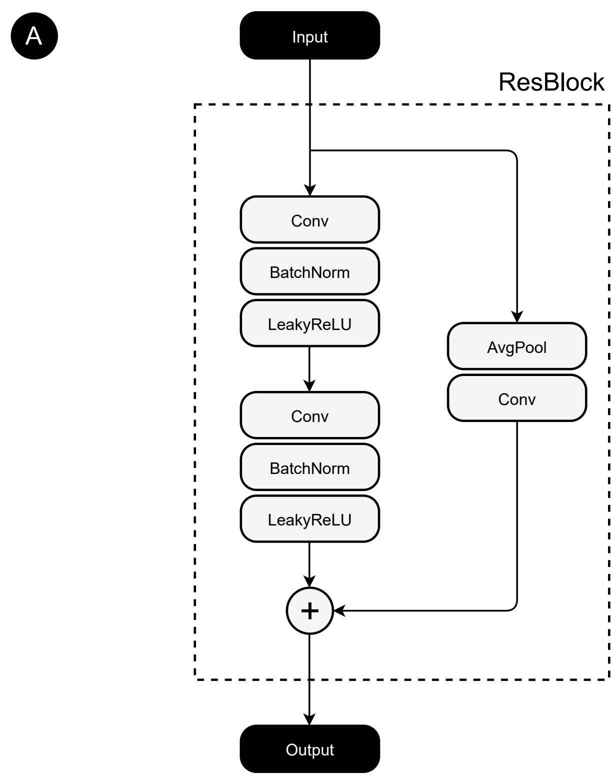
$$N = \sum_i X_i | Y_i \approx \sum_i X_i + Y_i - X_i * Y_i \quad (4)$$

where X_i and Y_i are outputs and targets for given class i, respectively. Since we are interested in maximizing the challenge score, we can invert the sign for challenge loss in eq. 1 and standardly use minimization optimizers.

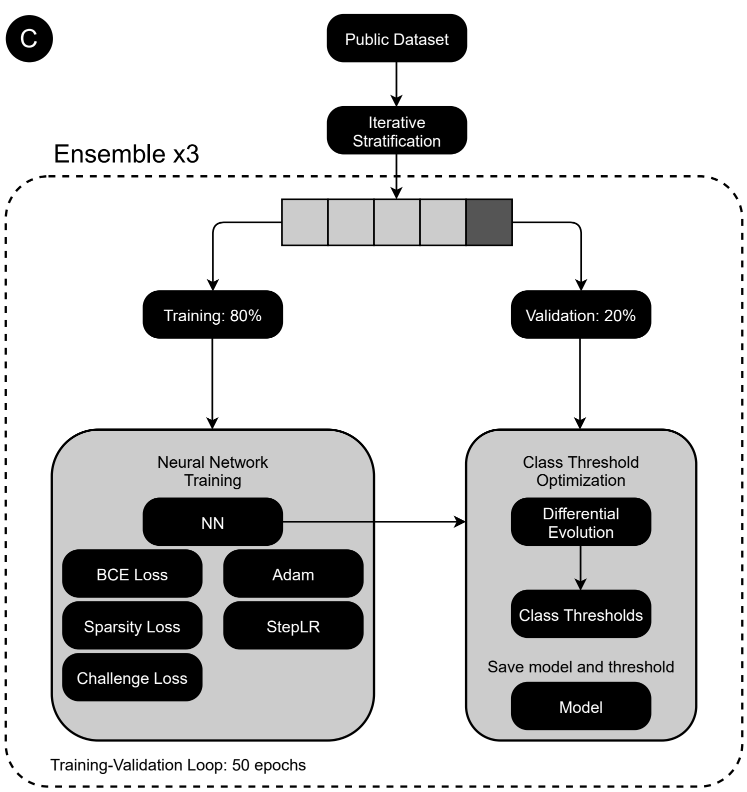
The model with the best validation performance is selected, and class thresholds are optimized by a differential evolution genetic algorithm. In general, this process requires a large amount of computation since we are exploring 26-dimensional space with boundaries between 0 to 1. The benefit of sparsity loss is that majority of model probability outputs are located close to 0 or 1, which speeds up the threshold optimization. The estimated class-specific thresholds are the same for all leads configurations. The random dataset split, model training, and threshold optimization are repeated three times to create the model ensemble. Each model in the ensemble outputs binary indicators for each class. For this reason, the final vote is decided by the majority vote.

3. Results

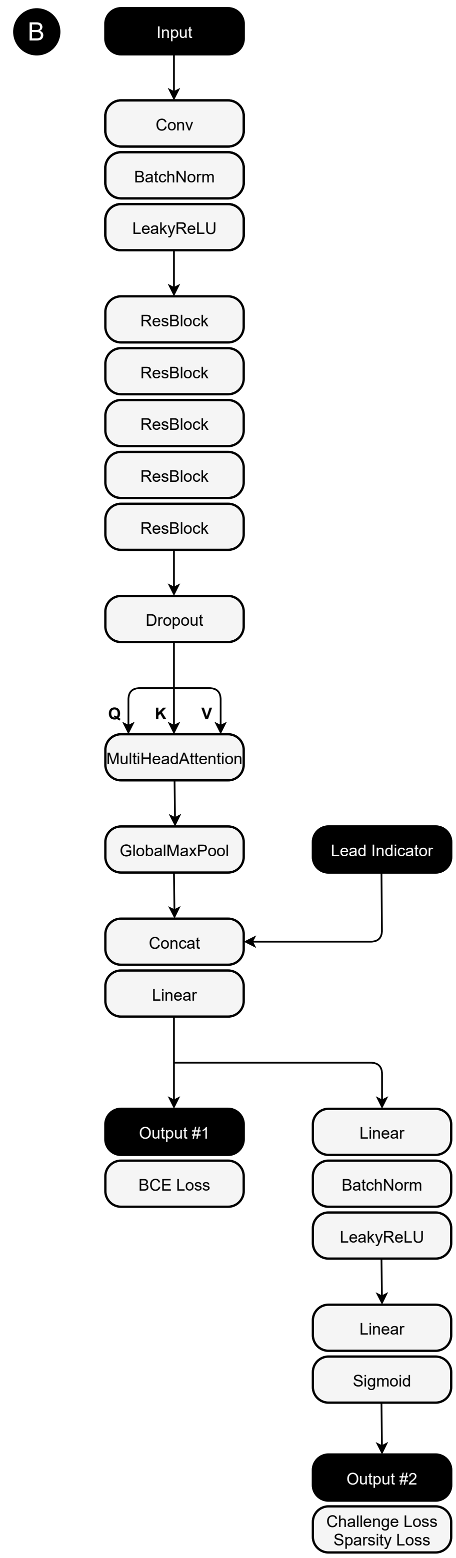
The performance of our algorithm (ISIBrno-AIMT team) was estimated on hidden validation set during the



(a) Residual block architecture.



(b) Training and validation pipeline with differential evolution threshold optimization.



(c) Full model architecture.

official phase of the challenge and obtained validation scores: 0.64, 0.62, 0.63, 0.63, and 0.62 for lead configurations 12, 6, 3, 4, and 2, respectively. The total training time was approximately 27 hours, i.e., 9 hours per model.

Local cross-validation results achieved a score of approximately 0.69, where only 12-lead performance was investigated. However, this result is biased since the local validation set is used for the selection of the best model and subsequent automatic threshold optimization.

Leads	Validation	Test	Ranking
12	0.64	???	???
6	0.62	???	???
4	0.63	???	???
3	0.63	???	???
2	0.62	???	???

Table 1: Challenge scores for our final selected entry (team ISIBrno-AIMT) scored on the hidden validation set, and one-time scoring on the hidden test set as well as the ranking on the hidden test set.

4. Discussion and Conclusions

This paper introduces our method for the classification of ECGs with a variable number of leads. We have developed a Residual CNN network with an attention mechanism that is optimized by a mixture of loss functions i.e., binary cross-entropy, differentiable approximation of challenge score, and sparsity loss function. Subsequently, a differential evolution algorithm is used for class-specific threshold optimization.

In comparison with our solution [4] from previous year challenge[8], we have improved preprocessing steps (filtering, normalization, and data augmentation) and slightly changed the architecture of the model (using large convolutional kernels). We hope that signal filtering in the narrow range of 1-47 Hz will help improve the generalizibility of our model, which was the critical drawback of our last year's solution.

Acknowledgments

This work was supported with a project by the Czech Academy of Sciences RVO:68081731 and by the Czech Technological Agency grant number FW01010305.

References

[1] WHO Cardiovascular diseases (CVDs). `https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)`. Accessed: 2021-08-30.

[2] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv 2015;.

[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] Nejedly P, Ivora A, Viscor I, Halamek J, Jurak P, Plesinger F. Utilization of Residual CNN-GRU With Attention Mechanism for Classification of 12-lead ECG. Computing in Cardiology 2020;1–4.

[5] Natarajan A, Chang Y, Mariani S, Rahman A, Boverman G, Vij S, et al. A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification. Computing in Cardiology 2020;1–4.

[6] Vicar T, Hejc J, Novotna P, Ronzhina M, Janousek O. ECG Abnormalities Recognition Using Convolutional Network With Global Skip Connections and Custom Loss Function. Computing in Cardiology 2020;1–4.

[7] Sechidis K, Tsoumakas G, Vlahavas I. On the Stratification of Multi-label Data. Machine Learning and Knowledge Discovery in Databases 2011;145–158.

[8] 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.

Address for correspondence:

Petr Nejedly
Královopolská 147, 612 00 Brno, Czech Republic
nejedly@isibrno.cz