

# Utilization of Residual CNN-GRU With Attention Mechanism for Classification of 12-lead ECG

Petr Nejedly, Adam Ivora, Ivo Viscor, Josef Halamek, Pavel Jurak, Filip Plesinger

Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

## Abstract

*Cardiac diseases are the most common cause of death. The fully automated classification of the electrocardiogram (ECG) supports early capturing of heart disorders, and, consequently, may help to get treatment early. Here in this paper, we introduce a deep neural network for human ECG classification into 24 independent groups, for example, atrial fibrillation, 1st degree AV block, Bundle branch blocks, premature contractions, changes in the ST segment, normal sinus rhythm, and others. The network architecture utilizes a convolutional neural network with residual blocks, bidirectional Gated Recurrent Units, and an attention mechanism. The algorithm was trained and validated on the public dataset proposed by the PhysioNet Challenge 2020. The trained algorithm was tested using a hidden test set during the official phase of the challenge and obtained the challenge validation score of 0.659 as entries by the ISIBrno team. The final testing scores were 0.847, 0.195, -0.006, and 0.122, for testing sets I, II, III, and full test set, respectively. We have obtained 30th place out of 41 teams in the official ranking.*

## 1. Introduction

Automated classification of ECG recordings has significant potential to improve current clinical practice by fast and accurate ECG interpretation. This might be especially helpful for the classification of long-term recordings, where data are continuously recorded over multiple days. In such cases, a human inspection of full data is nearly impossible due to time requirements and might be subjective. For this reason, an enormous scientific effort has been spent to automate classification pipelines and investigate various approaches mostly based on signal processing methods and machine learning techniques. Such approaches required manually designed ECG descriptors (features) that were subsequently

classified by machine learning algorithms such as logistic regression, SVM, feed-forward neural networks, or decision trees [1], [2]. However, this paradigm was recently replaced by deep-learning techniques using end-to-end learning, where useful features are automatically found during the learning process. Initially, the deep-learning methods were designed for image classification and subsequently adopted among various scientific fields like speech recognition, natural language processing, machine translation, and biological signal processing [3], [4].

Currently, the state of the art approaches for ECG classification utilize deep-learning techniques based on convolutional neural networks, recurrent neural networks, residual neural networks, attention mechanism, or its combinations of various kinds. Here in this paper, we propose a model based on Residual CNN blocks with a bidirectional GRU layer and an attention mechanism for multi-label 12-lead ECG classification into 24 independent groups solving PhysioNet Challenge 2020 [5].

## 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. Here, in this paper, we propose a custom ResNet [6] architecture that processes ECG data with sampling frequency 250 Hz. For this reason, the preprocessing pipeline utilizes polyphase filter resampling to standardize sampling frequency among



Figure 1. The picture depicts A) residual block architecture, B) full model architecture, and C) training and validation pipeline with grid-search threshold optimization

datasets. This sampling frequency was chosen based on our empirical experience. Moreover, in some exceptional cases, we allow direct processing of sampling frequencies close to 250 Hz, e.g., 257 Hz, which is incorporated in the publicly available training dataset. Each recording is converted to millivolts. Subsequently, each channel of ECG recording is normalized to zero mean (subtracting DC offset), which helps with model training.

To improve model performance, we have also included power envelopes in low (0-8 Hz, to amplify T-waves) and mid-frequency bands (0-24 Hz, to strengthen QRS complexes). The power envelope is evaluated as an absolute value of the Hilbert transform of the ECG signal filtered with an FFT filter (Tukey window function with  $\alpha = 0.5$ ).

The model architecture is based on residual CNN blocks, utilizing three 1x3 convolution kernels, with batch normalization and dropout regularization options. Each residual block down-samples the input signal by a factor of 2 by strided convolution. The full model is composed of five residual blocks, i.e. a total of 15 convolutional layers. Subsequently, extracted features from the last residual block are forwarded into the bidirectional GRU layer accompanied by an attention mechanism [7]. The final classification layer evaluates probabilities from estimated attention vectors for each classification class independently, i.e., utilizing the sigmoid activation function. The model is optimized by minimizing binary cross-entropy metrics by Adam optimization technique [8] with learning rate 1e-3 and L2 regularization parameter 1e-6 while reducing the learning rate by a factor of  $\frac{1}{2}$  when the validation score did not improve for seven epochs. The training minibatch was empirically chosen as 64 samples, where each recording was zero-padded into the length of 2 minutes (empirically determined).

For training and validation, the public challenge dataset was split into two sub-datasets in ratios 75% and 25%, respectively. The dataset stratification was iteratively optimized by method available in scikit-multilearn based on [9]. The model was trained for 75 epochs, while after each epoch, the validation dataset was used to evaluate challenge metrics score and optimize the global probability threshold by a grid-search method. During the model development, we have also evaluated the optimization of an individual threshold for each classification group by a differential evolution genetic

algorithm [10]. However, this did not improve the performance of our model. The model with the highest validation score was subsequently selected as the model for the inference phase.

### 3. Results

Entry	Validation Score	Description
1	0.569	Basic model with residual channels
2	0.624	Basic model with increased residual channels
3	0.608	Basic model with increased residual channels, without regularization and dropout
4	0.664	Power envelopes extended model with voltage normalization
5 (final)	0.659	Same as entry 4
6	0.632	Same as entry 2
7	0.652	Self-normalizing network

Table 1. Description of our challenge submissions. Validation score was obtained remotely from a hidden subset of challenge data

The entry 5 (Tab.1) was selected as the final entry and received a score of 0.847, 0.195, -0.006, and 0.122, for hidden testing sets I, II, III, and full test set, respectively.

### 4. Discussion

During the challenge, we have evaluated multiple input options and hyper-parameter settings. For the first entry, we have deployed the model with the proposed architecture and achieved the challenge score of 0.569. Subsequently, we have increased the number of channels in residual blocks to increase model performance, which improved the model performance to 0.624. In the next step, we tried to switch off the regularization and decrease

model dropouts. However, this did not yield a performance increase (0.608).

Subsequently, we have included the band-specific power envelopes. Simultaneously, we have changed the input normalization from z-score to voltage in mV since we observed the model performs poorly for the classification group “low QRS voltage”. Proposed changes increased the model performance to 0.664, which is the top result among our entries. Next, we re-evaluated entry 2 and 4, to test the numerical stability of the results, since model training and data split is a random process. Results suggest that outcomes are stable, where differences between entries are lower than 0.01 points of challenge score.

Our proposed method worked very well during the validation phase, however, the final challenge testing showed that the proposed algorithm is not generalizing well to the new data from other institutions. We suspect that the testing dataset exhibits a statistical distribution shift in comparison with training and validation. This might be observed by comparing testing scores from datasets I and III (0.847 vs -0.006), where data from III originates from a different hospital that is not included in the training. Moreover, at the moment, we are not able to address the loss of performance in dataset II (0.195). In summary, our model was probably overfitting towards one of the institutions in the training set. In further studies, we will investigate the causes of this overfitting issue.

## 5. Conclusion

In this paper, we have proposed a Residual-CNN GRU neural network with an attention mechanism for 12-lead ECG classification into 24 independent groups as the solution to the PhysioNet Challenge 2020. Our algorithm obtained 30th place out of 41 teams in the official ranking.

## Acknowledgments

The research was supported by a research grant no. FW01010305 by the Czech Technological Agency and by the Czech Academy of Sciences (project RVO:68081731).

## References

- [1] F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, “Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG,” *Physiol. Meas.*, vol. 39, no. 9, p. 094002, Sep. 2018.
- [2] F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, “Automatic detection of atrial fibrillation and other arrhythmias in Holter ECG recordings using PQRS morphology and rhythm features,” *2017 Computing in Cardiology Conference (CinC)*. 2017, doi: 10.22489/cinc.2017.364-057.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [4] Y. LeCun, D. Touresky, and G. Hinton, “A theoretical framework for back-propagation,” *Proceedings of the 1988 Connectionist Models Summer School*, 1988.
- [5] E. A. P. Alday *et al.*, “Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020,” *Physiol. Meas.*
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [7] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 5998–6008.
- [8] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” *arXiv*, Dec. 22, 2014.
- [9] K. Sechidis, G. Tsoumakas, and I. Vlahavas, “On the stratification of multi-label data,” in *Machine Learning and Knowledge Discovery in Databases*, 2011, pp. 145–158.
- [10] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

Address for correspondence:

Petr Nejedly

Ústav přístrojové techniky AV ČR, v. v. i.

Kralovopolska 147, Brno, Czech Republic

Email: [nejedly@isibrno.cz](mailto:nejedly@isibrno.cz)