Segment, Perceive and Classify – Multitask Learning of the Electrocardiogram in a Single Neural Network

Philipp F Sodmann[1,2,3], Marcus Vollmer[2,3], Lars Kaderali[2,3]

[1]Department of Internal Medicine I, University Hospital, Würzburg, Germany
[2]Institute of Bioinformatics, University Medicine Greifswald, Germany
[3]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Germany

Abstract

As part of the Physionet 2021 Challenge, "Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021", we have developed a neural network to classify pathologies and changes in the ECG. Our team HeartlyAI has developed a novel multitask learning based network that combines classification with segmentation and extrasystole detection. To obtain segmentation annotations, we developed an annotation tool in Angular and have manually annotated 1,789 ECGs from all challenge data sources for a gold standard of P wave, QRS, and T wave segments. Each extrasystole was annotated as supraventricular or ventricular.

In the first step of our classification workflow, the ECG is segmented using a U-Net. This segmentation is used to calculate within-net features such as the PQ, QTc time, and Q-Q interval. The bottleneck layer of the U-Net is used along with the computed features as an embedding for the classification. We have used the recent Perceiver architecture to perform the classification of the ECG.

Our classifiers received scores of 0.40, 0.31, 0.34, 0.34, and 0.25 (ranked 18th, 24th, 23rd, 23rd, and 27th) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the hidden validation set with the Challenge evaluation metric.

1. Introduction

The electrocardiogram (ECG) is an important tool in routine clinical practice for rapid diagnosis of life-threatening diseases and monitoring during interventions. Rapid and accurate detection of life-threatening pathologies in the ECG is critical to treat them at an early stage. This year's PhysioNet/Computing in Cardiology Challenge aims to detect pathological changes in ECGs with a reduced set of leads [1–3].

In total, more than 88,253 ECGs [4–9] were provided for training purposes along with 29 clinical diagnoses, of which 3 have been grouped together.

Our approach follows the trend of end-to-end trained neural networks and multitask learning. Currently, the Transformer architecture is established in more and more areas and is replacing convolutional neural networks [10]. However, their quadratic memory requirements are challenging for long ECG sequences that can exceed half an hour. There are several approaches to solve this, one is to predict on shorter parts of the ECG and then merge them back later. The second is to bring the shape within the network into a more condensed form. We used the latter together with the Perceiver [11], where the query vector is separated from the input, making it have only a linear memory complexity.

2. Methods

Our network consists of three main components. A U-Net for segmentation of the ECG into P-wave, QRS complex and T-wave and a classification head for classification of the QRS complexes into normal beats, ventricular and supraventricular extrasystoles. Based on this, a module that calculates the PQ time, the Q-Q distance and the QT distance as well as the Fridericia QTc time within the network. These times were repeated for the corresponding heartbeat. The bottleneck layer of the U-Net is concatenated together with the times and serves as an embedding of the ECG reduced by a factor of 64 in temporal dimension. This classification branch of the network consists of a Perceiver [11] module with iterative attention with 256 outputs and a linear layer to reduce to the 29 classes.

During preprocessing, strong signal peaks were first removed and the signal interpolated if the difference between the previous and the following frame was greater than $4\,\mathrm{mV}$. A Butterworth filter with a cut-off frequency of $50\,\mathrm{Hz}$ and an edge frequency of $60\,\mathrm{Hz}$ was then applied. Locally weighted scatterplot smoothing was then applied to remove the baseline wander. Finally, the trimmed mean was subtracted to center the signal around zero. Lastly, the

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.129

ECG voltage was amplified when the 98 percent quantile was below $1\,\mathrm{mV}$.
Since all ECGs needed the same length for the training process, segments with a length of 4096 frames were selected at random. If a manual segmentation was available for the respective ECG, the section was selected so that it was within the annotated 20 seconds. Shorter ECGs were zero-padded to a length of 4096. The entire length of the ECG was used at once for the runtime prediction. Training sequences longer than 4096 frames resulted in higher computational cost but did not add value to training in any metric.

To make classification easier for the network, we have modified the diagnostics within the network. We have split the diagnosis ”Normal sinus rhythm” into ”healthy” and ”sinus rhythm” and added a category ”pathological”. Each ECG with ”normal sinus rhythm” got the labels ”sinus rhythm” and ”healthy”, while all ECGs without a rhythm diagnosis (including those not evaluated in the challenge) got the labels ”sinus rhythm” and ”pathological”. Furthermore, we divided ”sinus bradycardia” into the diagnoses ”sinus rhythm”, ”bradycardia”, and ”pathological”.

We randomly divided all data into 10 bins and used 8 of them for pre-training and training, one for validation and the last for testing. This also allows for 9-fold cross-validation.

To pre-train our model, we disabled the linear projection layer of the classification layer and trained the classification head self-supervised with redundancy reduction [12]. The two other heads of the model were trained normally. This guaranteed that the segmentation in the final model was already well converged, and a good distribution for the embedding was already learned for the classification. The pre-training was performed on our training split of the challenge data.

The encoder of the U-Net consists of residual blocks with two convolutions with a filter size of five. For downsampling, we used max pooling with filter size and stride of two. The number of filters was not further tuned, in Table 1 the number of filters and the resulting temporal reduction are listed. In the decoder, the features were upscaled with a transposed convolution, the skip connections were concatenated, and two residual convolutions were applied. To guarantee that the features run through the whole U-Net, skip connections were limited to only 8 channels. Furthermore, deep supervision was used to obtain a prediction of the segmentation with reduced resolution at each level of the U-Net. Originally, Mish [13] was used as the activation function, but due to problems with older PyTorch versions used on the challenge servers, it was replaced by SiLU activation, which converged slightly worse, but did better than ReLU.

Encoder filters	Temporal reduction	Decoder filters
32	1	32
64	2	24
112	4	48
192	8	64
256	16	96
320	32	128
512	64	

Table 1. Number of filters in the U-Net encoder.

We used the Tversky Loss with alpha=0.7 and beta=0.3 for segmentation, Binary Crossentropy for beat classification and Asymmetric Loss for pathology classification with gamma_neg=2, gamma_pos=1 and clip=0.05. To weight the individual loss functions, we multiplied the segmentation loss by 4 and the beat classification loss by 2.

As the Madgrad [14] optimizer converged faster than Adam [15] and Ranger [**?**], we used Ranger with an Asymmetric Loss For Multi-Label Classification (ASL) [16], which outperformed binary cross-entropy and focal loss. The ASL was used with a negative gamma of 2 and a positive gamma of 1 which penalized false negatives more severely than false positives during training and led to higher recall than precision. The model was trained for 180 epochs, each epoch containing all manually segmented ECGs as well as 12,000 additionally sampled ECGs. For sampling, the probability for each ECG to be selected was inversely proportional to the frequency of its positive classes. After 40, 80, and 140 epochs, the gradient accumulation was increased quadratically. We used a cyclic learning rate scheduler with a maximum learning rate of 2e-5.

We used different augmentation techniques such as cutout, adding different types of noise, as well as dropout of individual ECGs and groups of ECG channels. Due to the long training time, we only trained one model that can handle different combination of ECG channels. The overall network architecture is illustrated in Figure 1.

An interactive visualization of the segmentation can be tested online: `https://heartly.ml`

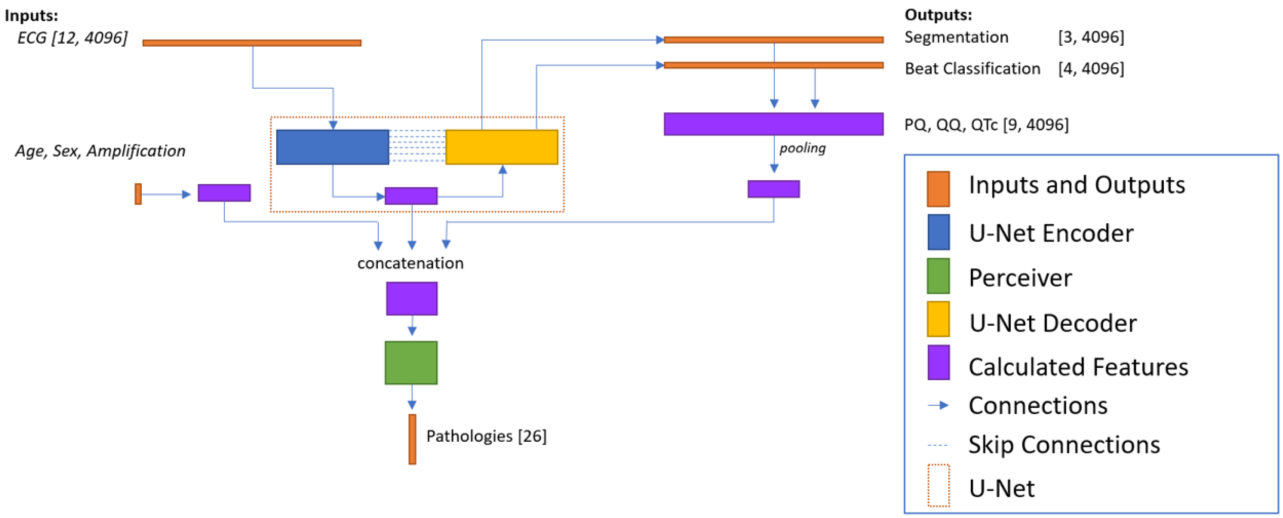

Figure 1. Overview of the network design. The network consists of two inputs, one for the ECG and one for the age, gender and 95 percent quantile of the ECG (amplitude). It has 3 outputs: segmentation, beat classification and pathology classification. The bottleneck of the U-Net together with the age, sex and amplitude as well as features of the segmentation and beat classification are used for the classification of the whole ECG.

,

3. Results

The performance on our validation split was very good for some classes, "Pacing Rhythm", "right bundle branch block" as well as "sinus bradycardia" and "sinus tachycardia" all achieved a F1 score of more than 0.9. However, other classes, such as bundle branch blocks and right axis deviation, only had an F1 score of less than 0.4. Precision and recall were relatively balanced, with a slight emphasis on recall.
Segmentation reached dice values of 0.77 for P waves, 0.92 for T waves, and 0.97 for QRS complexes on our validation data. Unfortunately, no metrics were calculated for beat classification

Leads	Training	Validation	Test	Ranking
12	0.59	0.44	0.4	18
6	0.39	0.35	0.31	24
4	0.52	0.40	0.34	23
3	0.50	0.40	0.34	23
2	0.38	0.32	0.25	27

Table 2. Challenge scores calculated on a 10% test split of the public training set, repeated scoring 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

For this challenge we have developed a model that can combine various tasks such as segmentation, beat classification and pathology classification in a single model. Due to a restricted computational budget, we were not able to investigate the hyperparameter space for our model. During training, the model was able to overfit and correctly classified almost all ECGs. It achieved a micro F1 Score of 0.98 while the validation split did not improve beyond 0.82. Particularly, the results for 6 leads were worse than the other combination due to a bug in our augmentation step, which was only discovered after the challenge had ended. We also found an error that caused the model to fail in predicting the abnormal and inverted T-wave classes. After solving the latter problem, our challenge score on our private validation split increased to 0.65 for 12-channel ECGs. Optimization of the hyperparameters could further help to increase the performance score. Our team Heartly AI was able to show that multitask learning can be successfully applied to ECGs.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[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] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. CoRR 2020;abs/2010.11929. URL https://arxiv.org/abs/2010.11929.

[11] Jaegle A, Gimeno F, Brock A, Zisserman A, Vinyals O, Carreira J. Perceiver: General perception with iterative attention. CoRR 2021;abs/2103.03206. URL https://arxiv.org/abs/2103.03206.

[12] Zbontar J, Jing L, Misra I, LeCun YA, Deny S. Barlow twins: Self-supervised learning via redundancy reduction. In ICML. 2021; .

[13] Misra D. Mish: A self regularized non-monotonic neural activation function. ArXiv 2019;abs/1908.08681.

[14] Defazio A, Jelassi S. Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization. arXiv preprint arXiv210111075 2021;.

[15] Kingma DP, Ba J. Adam: A method for stochastic optimization. CoRR 2015;abs/1412.6980.

[16] Baruch EB, Ridnik T, Zamir N, Noy A, Friedman I, Protter M, et al. Asymmetric loss for multi-label classification. ArXiv 2020;abs/2009.14119.

Address for correspondence:

Philipp F Sodmann
Am Exerzierplatz 3a, 97072 Würzburg, Bavaria, Germany
sodmann_p@ukw.de