ECG Segmentation Using a Neural Network as the Basis for Detection of Cardiac Pathologies

Philipp Sodmann[1, 2], Marcus Vollmer[1, 2]

[1]Institute of Bioinformatics, University Medicine Greifswald, Germany
[2]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Germany

Abstract

Electrocardiography allows fast and noninvasive diagnosis and screening of a wide range of cardiac diseases. The interpretation of ECGs is difficult and depends on the levels of training of the physician. In consequence, pathologies can remain undiagnosed or norm-variations are interpreted as pathological.

The PhysioNet/Computing in Cardiology Challenge 2020 aims to classify various cardiac pathologies in 12-lead ECGs, data was collected across a variety of different clinics and countries to pave the way for a common evaluation of ECGs. Our Team Heartly-AI proposes a two-step algorithm using a UNet and XGBoost for the 2020 PhysioNet Computing in Cardiology Challenge "Classification of 12 lead ECGs". The algorithm achieved a 5-fold cross-validation metric of 0.113 and scored 0.136 on the official test set, therefore placing us 28th out of the 41 teams in the official ranking.

1. Introduction

Electrocardiography allows fast and noninvasive diagnosis and screening of a wide range of cardiac diseases ranging from arrhythmias like atrial fibrillation, conduction disorders like branch blocks to myocardial infarction [1, 2]. The interpretation of ECGs is difficult and depends on the level of training of the physician [3]. General practitioners (GPs), who often record ECGs, have a diagnostic agreement of around 70% [4–6]. Automatic interpretation performed by ECG machines have shown an even worse performance around 44% agreement [6, 7] In consequence, pathologies can remain undiagnosed or norm-variations are interpreted as pathological. In the last years, there has been a lot of active research on Deep Neural Networks for ECG analysis, but they often only focus on a few common diagnoses and the data originates from only one clinic [8–10]. This makes it all the more important that the focus should be on identifying numerous different diagnoses in ECGs coming from a diverse range of data sources.

Our approach to this year's PhysioNet Challenge [11, 12] was to develop a solution that creates human-interpretable intermediate outputs and thereby making the models' classification understandable. First, we segmented the ECG into P waves, QRS complexes, T waves, and interbeat segments with an UNet-like fully convolutional neural network. In the second step, we calculated features like median PQ time or R-amplitude. Lastly, we used XGBoost to classify the ECG with 24 non-exclusive labels.

2. Methods

2.1 Preprocessing

For preprocessing, the ECGs with a varying sampling frequency were resampled to $500\,\text{Hz}$ using the Fast Fourier transform. To remove baseline wander and high-frequency noise, the resampled signal was filtered with a band-pass filter to a range of $0.05\,\text{Hz}$ to $42\,\text{Hz}$. Furthermore, we applied a discrete wavelet transform with Daubechies wavelet (Db4) with hard thresholding as a high-pass filter to remove the remaining baseline wander.

2.2 Training Data

Training data for the segmentation was generated by building an annotation tool that allowed us to mark segments in ECGs as either P waves, QRS complexes, or T waves. The QRS complexes were annotated as either normal, supraventricular, or ventricular extrasystoles. With this tool, we annotated snippets of up to 20 seconds of random ECGs. The backend of the annotation tool was programmed in Python. FastAPI was used to preprocess and provide the ECG files and store the data in a MySQL database. The annotation frontend was built with Angular as a web app. To display the data, ECGs were drawn on a canvas object with a resolution of 1 pixel per data point (see Figure 1). During annotation, the frontend displayed predictions for the segmentation that could easily be adjusted or completely removed. All data was annotated by one physician. As a labeling strategy, we first manually annotated 200 random ECG snippets. Second, we used
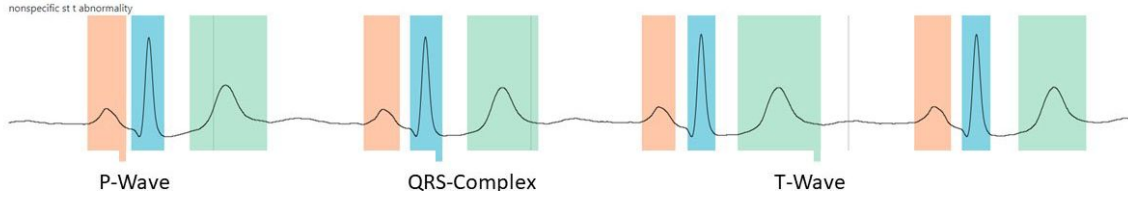
Figure 1. Example annotation of an ECG taken from our annotation tool. P-waves are highlighted in orange, QRS-complex in blue and T-waves in green. All the QRS-complexes are annotated as normal beats.

a neural network (see Figure 2) for automated prelabeled segmentation of 200 additional ECGs. These were manually screened and segments were corrected if necessary. The rules for selecting cases were built on physiological plausibility, for example, finding a T-wave after a QRS-complex vs. two QRS-complexes occurred in succession. Moreover, we screened and corrected another 330 ECG records in which there was a discrepancy between labeled extrasystoles and the prediction of our neural network. Because calibration bars in records resulted in artifacts, we also added 10 annotated ECGs containing these artifacts to the annotated corpus. Each annotated ECG was randomly assigned into one of 10 groups. This allowed us to maintain training, validation, and test splitting throughout the annotation process and to keep track of the performance progress. We trained the network on 573, validated on 94, and tested on 73 ECGs.

2.3   Training the UNet

The trained neural network was inspired by an existing solution for ECG segmentation [13]. We trained the network on 12-lead ECGs of arbitrary length with all 12 channels as inputs, and two different outputs with four neurons each. The first output predicted for each time point whether it belonged to a P-wave, T-wave, or QRS-complex. The second output predicted for each QRS-complex if it is a supraventricular or ventricular extrasystole, a normal beat, or no QRS-complex (see Figure 2).

The input of the network accepts a signal with any number of time steps and 12 channels (None, 12), allowing us to pass ECGs of any length. We down-sampled the signal 4 times with blocks of 2 convolutions and a max-pooling with a stride of 2. Thus, the time dimension of the middle layers were 1/16th of the original duration. They consisted of 2 convolutional layers, each with a dropout of $50\,\%$. To upsample the network to the original duration, each step consisted of one transposed convolution, doubling the time dimension, and three dilated convolutions. The down-sampling and upsampling blocks were summed to create skip connections. The primary output for the segmentation are 3 convolutional layers and one time-distributed dense layer, which is equal to 4 convolutions with size 1. For extrasystole classification, there is a secondary output with 2 convolutional layers, and a dense layer with softmax multiplied with the QRS output.

For each layer, we tried a set of 32, 48, 64, 72, 80, 88, 96 neurons while restricting previous layers to the same amount or fewer neurons than in the previous layer. We choose the amount of neurons that was the smallest without worsening the prediction score.

The applied losses were Tversky loss for the segmentation [14] and categorical cross-entropy for the extrasystole classification. For segmentation, we also tried other losses like dice loss[15], focal Tversky, and Tanimoto loss [16] and categorical cross-entropy, but they did not achieve proper segmentation boundaries.
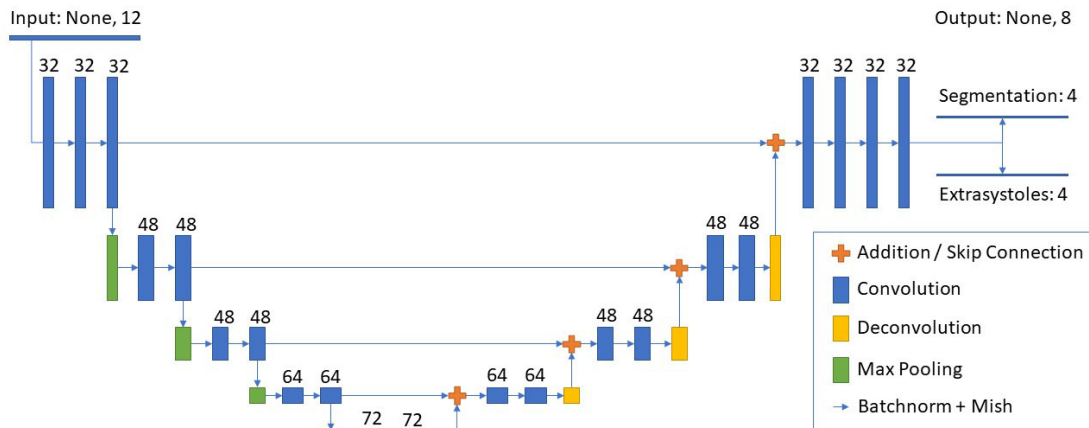

Figure 2. Scheme of our UNet like convolutionary neural network.

In general, each convolution was followed by batch normalization and Mish activation [17]. As an optimizer, Rectified Adam with LookAhead (Ranger) was used [18]. The network was trained for 200 epochs. After 140 epochs, we reduced the learning rate with a cosine annealing to zero [19]. This way, the network showed minimal over-fitting when training on the training data. In our official entries, we trained with the same parameters on the whole annotated dataset.

2.4 Segmentation

Using our UNet, we were able to calculate to approximate the onsets and offsets of the ECG segments. Each QRS-complex was part of a beat object which also contained the corresponding P-wave and T-wave. We further validated each heartbeat: A beat object must contain a QRS-complex and a T-wave to be valid. Invalid beats as well as the first and last beat were excluded from all feature calculations. We tried more strict rules such as a minimum length of QRS-complex and T-wave but this did not improve the classification performance. Additionally, we estimated the baseline of the ECG signal by linear regression of interbeat intervals between each beat. We used this baseline to calculate the amplitudes of P and T-waves as well as the QRS-complexes.

2.5 Feature extraction

We extracted 54 features from the segmented ECGs. Classical features such as heart rate, QRS complex amplitude, PR interval, and QTc time were calculated from the segmented data. For each measured duration, mean and median absolute deviation were extracted. We computed the Frank leads [20] and calculated the amplitude in the X, Y, and Z dimension and the angle in the XY, YZ, and XZ plane for P, R, and T-waves. Extrasystole information was used to calculate the relative proportion of normal beats as well as supra and ventricular extrasystoles. To distinguish left and right branch blocks, we calculated the time difference of the R peak in V1, V2 and V5, V6.

2.6 Classification

To tune the parameters of XGBoost, we used 5-fold cross-validation [21] using hyperopts implementation of tree-structured Parzan estimators [22] (see Table 2). Tuning objective was to maximize the prediction accuracy.

Lastly, a model was trained on all data using the optimized set of parameters. Due to time constraints, we did not perform any relabeling of the training data.

Table 1. R-Peak segmentation accuracy in F1 on the MIT-BIH Database (MITDB).

Tolerance	5 ms	10 ms	20 ms	30 ms	150 ms
MITDB	0.595	0.880	0.922	0.932	0.943

Table 2. Best set of parameters using hyperparameter optimization of the XGBoost model.

Parameter	Default	Tuned
n_estimators	100	300
max_depth	6	7
gamma	0	0.9816
alpha	0	0.9816
min_child_weight	1	1
subsample	1	0.77788
learning_rate	0.3	0.06167

Table 3. Segmentation accuracy on a test split (F1).

Tolerance	P-wave	QRS-complex	T-wave
5 ms	0.7765	0.9210	0.5967
10 ms	0.9575	0.9790	0.7003
20 ms	0.9855	0.9816	0.8834
30 ms	0.9855	0.9816	0.9225

3. Results

While the segmentation showed promising results, our model was unable to achieve an overall good performance. The model achieved a challenge metric of 0.136 on the hidden test set while the top-performing teams scored up to 0.533. This ranked our team Heartly-AI on the 28th place of the participating 41 teams.

On a validation split of the data, only 8 of the scored 24 classes reached an F1 greater than 0.5. The model performed well on atrial fibrillation (F1 = 0.87), sinus rhythm (F1 = 0.82) and left bundle branch block (F1 = 0.77).

We compared the signal onset with the predicted onset given a certain tolerance to test the segmentation accuracy on a test split, see Table 3. Table 1 shows the performance on the MIT-BIH Database. To match our data format, we repeated each of the two channels in the MITDB 6 times to simulate 12 channels. This might lead to a worse performance compared with genuine 12 channels.

4. Discussion and Conclusions

We developed and tested an algorithm for segmentation and classification of 12-lead ECGs. Even though our approach for segmentation did not achieve a state of the art result in beat detection, its performance is comparable to other deep learning based approaches [23–25].

The classification part of our algorithm did not perform satisfactorily. One problem with our algorithm was that the provided labels were only annotated for 6 classes in the CPSC dataset. However, we trained on 24 classes, making the correct classification of the other 18 classes even less likely. We also prepared an entry that masked out the labels for not annotated categories in the CPSC data, but this did also affect the validation error since we created false positive matches in non-annotated classes and thus resulted in

a worse challenge metric. We did not include the source of the data in the features since this would contradict the purpose of the challenge to build a classifier for any source of data. Further, it would not generalize to new unseen databases that will be in the final test data. Moreover, we did not correct any labels of incorrectly classified ECGs.

Our original plan was to use the segmentation data as an auxiliary output for a network that would classify the pathology in the ECG. In preliminary tests on our own 740 annotated ECGs, our approach has shown promising results, achieving a macro F1 of 0.5. However, this plan was no longer pursued due to a lack of time to implement this methodology in a way that it would reliably train on the PhysioNet servers, and we hope to test this solution after the challenge. We also see potential in annotating further ECGs with active learning. For example, we could focus on ECGs with first-degree AV-blocks where the $P - R$ distance is detected in normal ranges. Also, a wider variety of training sources could help to improve the segmentation task even further.

Acknowledgments

References

[1] German DM, Kabir MM, Dewland TA, et al. Atrial fibrillation predictors: Importance of the electrocardiogram. Annals of Noninvasive Electrocardiology 2016;21(1):20–29.

[2] Birnbaum Y, Wilson JM, Fiol M, et al. ECG diagnosis and classification of acute coronary syndromes. Ann Noninvasive Electrocardiol 2014;19(1):4–14.

[3] Massel D. Observer variability in ECG interpretation for thrombolysis eligibility: Experience and context matter. J Thormb Thrombolys 2003;15(3):131–140.

[4] Santos P, Pessanha P, Viana M, et al. Accuracy of general practitioners readings of ecg in primary care. Central European Journal of Medicine 2014;9:431–436.

[5] Nieuwenhof N, Willemsen R, Konings K, et al. Interpretations of and management actions following ECGs in programmatic cardiovascular care in primary care: A retrospective dossier study. Netherlands Heart Journal 2020;28.

[6] Whitman M, Layt D, Yelland M. Key findings on ECGs - level of agreement between GPs and cardiologists. Aust Fam Physician 2012;41:59–62.

[7] Schläpfer J, Wellens HJ. Computer-interpreted electrocardiograms: Benefits and limitations. Journal of the American College of Cardiology 2017;70(9):1183–1192.

[8] Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020;11(1):1760.

[9] Smith SW, Rapin J, Li J, et al. A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. IJC Heart Vasculature 2019; 25:100423.

[10] Kim JH, Seo SY, Song CG, Kim KS. Assessment of electrocardiogram rhythms by GoogLeNet deep neural network architecture. Journal of Healthcare Engineering 2019; 2019:2826901.

[11] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[12] Perez Alday EA, Gu A, Shah A, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol Meas 2020;.

[13] Moskalenko V, Zolotykh N, Osipov G. Deep learning for ECG segmentation. In Studies in Computational Intelligence. Springer International Publishing, 2019; 246–254.

[14] Abraham N, Khan NM. A novel focal tversky loss function with improved attention U-Net for lesion segmentation 2018;arXiv:1810.07842.

[15] Sudre CH, Li W, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations 2017;arXiv:1707.03237.

[16] Diakogiannis FI, Waldner F, Caccetta P, Wu C. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data 2019;arXiv:1904.00592.

[17] Misra D. Mish: A self regularized non-monotonic activation function 2019;arXiv:1908.08681.

[18] Yong H, Huang J, Hua X, Zhang L. Gradient centralization: A new optimization technique for deep neural networks 2020;arXiv:2004.01461.

[19] Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts 2016;arXiv:1608.03983.

[20] Daniel G, Lissa G, Redondo DM, et al. Real-time 3d vectorcardiography: an application for didactic use. Journal of Physics Conference Series 2007;90:012013.

[21] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. ACM, 2016; 785–794.

[22] Bergstra J, Yamins D, Cox DD. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In JMLR Workshop and Conference Proceedings, volume 28. 2013; 115–123.

[23] Jimenez-Perez G, Alcaine A, Camara O. U-Net architecture for the automatic detection and delineation of the electrocardiogram. In Computing in Cardiology. 2019; 1–4.

[24] Malali A, Hiriyannaiah S, G.M. S, et al. Supervised ECG wave segmentation using convolutional LSTM. ICT Express 2020;6(3):166–169.

[25] Sereda I, Alekseev S, Koneva A, et al. ECG segmentation by neural networks: Errors and correction. In 2019 International Joint Conference on Neural Networks. 2019; 1–7.

Address for correspondence:

Philipp Sodmann,   sodmann_p@ukw.de

Am Exerzierplatz 3a, 97072 Würzburg, Germany