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 [1, 2], 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". We scored 0.159 on the official testset and ranked about 199th out of the 200 teams that participated in this year's Challenge.

1. Introduction

Electrocardiography allows fast and noninvasive diagnosis and screening of a wide range of cardiac diseases (example...). The interpretation of ECGs is difficult, and depends on the levels of training of the physician [3]. General practitioners (GPs), who often record ECGs, have a diagnositc 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 a large number of different diagnoses in ECGs coming from a diverse range of data sources.

Our approach to this years physionet challenge 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 $P-Q$ time or R-amplitude. Lastly, we used XGBoost to classify the ECG with 24 non exclusive labels.

2. Methods

1. Preprocessing

For preprocessing, the ECGs with a different sampling frequency were resampled to 500 Hz using a fast fourier transform. To remove baseline wander and high frequency noise, the resampled signal was filtered with a low and high bandpass filter to a range of 0.05 Hz to 42 Hz. Furthermore, we applied a discrete wavelet transform with Debaucheries Wavelet (Db4) with hard thresholding as a high-pass filter to remove remaining baseline wander.

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 different ECGs. The backend of the annotation tool was programmed in Python. FastAPI was used to preprocess and serve 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 (Fig.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 annotated 200 random ECGs to bootstrap the annotation. With a rule based active learning approach, we annotated 200 ECGs. The
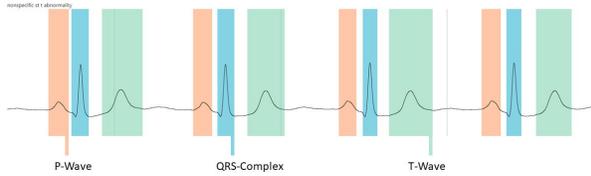
Figure 1. Example annotation of an ECG taken from our annotation tool. P-waves are highlighted in orange, QRS-complex in blue and T-waves in green. All of the QRS-complexes are annotated as normal beats.

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

rules selected cases where, for example, the predictions were physiological incorrect, we did not find a T-wave after a QRS-complex or two QRS-complexes occurred in succession. Moreover, we annotated another 330 ECGs 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 keep fixed training, validation and test split throughout the annotation phase and track our performance progress.

3. Training the UNet

The trained neural network was inspired by an existing solution for ECG segmentation [11]. We trained the network on ECGs of arbitrary length with 12 channels as inputs, and two different outputs with four neurons each. The first output predicted whether a certrain timestep 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 (Fig.2)

The input of the network accepted a signal with a shape of (None, 12), allowing us to pass ECGs of variable 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 downsampling 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
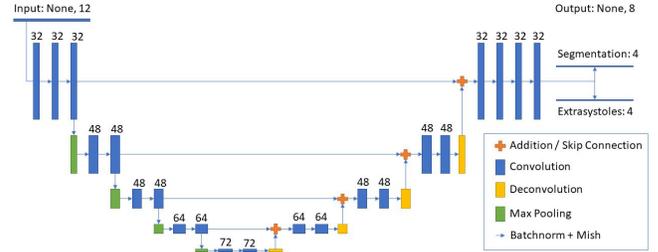
amount or less neurons than in the previous layer. We chose the amount of neurons that was the smallest without worsening the prediction score.

The applied losses were tversky loss for the segmentation [12] and categorical-crossentropy for the extrasystole classification. For segmentation, we also tried other losses like dice loss[13], focal tversky and tanimoto loss [14] and categorical crossentropy but they did not achieve proper segmentation boundaries.

In general, each convolution was followed by batch normalization and mish activation [15]. As an optimizer, Rectified Adam with LookAhead (Ranger) was chosen [16]. The network was trained for 200 epochs. After 140 epochs, we reduced the learning rate with a cosine annealing to zero [17]. This way, the Network showed minimal overfitting when training on the training data. In our official entries, we trained with the same parameters on the whole annotated dataset.

4. Segmentation

Using our UNet, we were able to calculate an estimation of the ECG segmentation. 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 performance. Additionally, we estimated the baseline by linear regression of $T - P$ intervals for each beat.

5. Feature extraction

We extracted 54 features from the segmented ECGs. Classical features such as heartrate, QRS complex amplitude, PR interval and QTc time were calculated from the segmented data. For each measured duration, we calculated mean and median absolute deviation. We computed the Frank leads [18] 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-Wave. The extra systole information was used to calculate the relative proportion of

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 1. Best found set of parameters using hyperparameter optimization of the XGBoost model.

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.

6. Classification

To tune the parameters of XGBoost, we used a 5 fold crossvalidation [19] using hyperopts [20] implementation of tree of parzan estimators [21] [Table 1]. The objective of the tuning was to maximize the accuracy of the prediction.

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.

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 0.159 while the top performing teams scored up to 0.696.

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

We also tested the segmentation against the MIT-BIH Database (MITDB). To match our dataformat, 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.

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

Segmentation accuracy in F1 on the MITDB.

4.   Discussion and Conclusions

We developed and tested an algorithm for segmentation and classification of 12 channel ECGs. Even though our algorithm for segmentation did not achieve state of the art

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

Table 2. Segmentation accuracy in F1 on a test split of our segmentation data. We compared the signal onset with the predicted onset given a certain tolerance.

results in beat detection, its performance is comparable other deep learning based approaches for ECG segmentation [22–24].

The classification part of our algorithm did not perform as well. One issue 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 effect the validation error since we created false positive matches in the not 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 740 annotated ECGs, this 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] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. 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, Liu F, Rad BA, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Under Review 2020;.

[3] Massel D. Observer variability in ECG interpretation for thrombolysis eligibility: Experience and context matter. Journal of Thrombosis and Thrombolysis June 2003; 15(3):131–140. URL https://doi.org/10.1023/b:thro.0000011368.55165.97.

[4] Santos P, Pessanha P, Viana M, Campelo M, Nunes J, Hespanhol A, Macedo F, Couto L. Accuracy of general practitioners' readings of ecg in primary care. Central European Journal of Medicine 06 2014;9:431–436.

[5] Nieuwenhof N, Willemsen R, Konings K, Stoffers H. Interpretations of and management actions following ecgs in programmatic cardiovascular care in primary care: A retrospective dossier study. Netherlands Heart Journal 02 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(1-2):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. ISSN 0735-1097. URL http://www.sciencedirect.com/science/article/pii/S0735109717387946.

[8] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira Jr. W, Schön TB, Ribeiro ALP. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature Communications Apr 2020;11(1):1760. ISSN 2041-1723. URL https://doi.org/10.1038/s41467-020-15432-4.

[9] Smith SW, Rapin J, Li J, Fleureau Y, Fennell W, Walsh BM, Rosier A, Fiorina L, Gardella C. 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. ISSN 2352-9067. URL http://www.sciencedirect.com/science/article/pii/S2352906719301241.

[10] Kim JH, Seo SY, Song CG, Kim KS. Assessment of electrocardiogram rhythms by googlenet deep neural network architecture. Journal of Healthcare Engineering Apr 2019; 2019:2826901. ISSN 2040-2295.

[11] Moskalenko V, Zolotykh N, Osipov G. Deep learning for ECG segmentation. In Studies in Computational Intelligence. Springer International Publishing, September 2019; 246–254. URL https://doi.org/10.1007/978-3-030-30425-6_29.

[12] Abraham N, Khan NM. A novel focal tversky loss function with improved attention u-net for lesion segmentation. arXiv preprint arXiv181007842 2018;.

[13] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, 2017.

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

[15] Misra D. Mish: A self regularized non-monotonic activation function, 2019.

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

[17] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts, 2016.

[18] Daniel G, Lissa G, Redondo DM, Vásquez L, Zapata D. Real-time 3d vectorcardiography: an application for didactic use. Journal of Physics Conference Series nov 2007; 90:012013. URL https://doi.org/10.1088%2F1742-6596%2F90%2F1%2F012013.

[19] 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. New York, NY, USA: ACM. ISBN 978-1-4503-4232-2, 2016; 785–794. URL http://doi.acm.org/10.1145/2939672.2939785.

[20] Bergstra J, Yamins D, Cox D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In International conference on machine learning. 2013; 115–123.

[21] Bergstra J, Yamins D, Cox DD. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ICML'13. JMLR.org, 2013; I–115–I–123.

[22] Jimenez-Perez G, Alcaine A, Camara O. U-net architecture for the automatic detection and delineation of the electrocardiogram. In 2019 Computing in Cardiology (CinC). 2019; Page 1–Page 4.

[23] Malali A, Hiriyannaiah S, G.M. S, K.G. S, N.T. S. Supervised ecg wave segmentation using convolutional lstm. ICT Express 2020;6(3):166 – 169. ISSN 2405-9595. URL http://www.sciencedirect.com/science/article/pii/S2405959520300989.

[24] Sereda I, Alekseev S, Koneva A, Kataev R, Osipov G. Ecg segmentation by neural networks: Errors and correction. In 2019 International Joint Conference on Neural Networks (IJCNN). 2019; 1–7.

Address for correspondence:

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