Pathologies Prediction on Short ECG Signals with Focus on Feature Extraction Based on Beat Morphology and Image Deformation

Jeffrey van Prehn[1], Svetoslav Ivanov[1], Georgi Nalbantov[1]

[1]Data Science Consulting Ltd, Pernik, Bulgaria

Abstract

Automated detection of key cardiac pathologies in reduced-lead ECGs is an enabling factor in applying ECG analysis on a larger scale. The PhysioNet/Computing in Cardiology Challenge 2021 identifies a set of key cardiac pathologies and challenges us with the task to automatically detect them. Critical to this task is the extraction of features from these ECGs which, combined, mark the presence of one or more of these key cardiac pathologies.

Methodology: algorithms were devised to automatically extract features based on the definitions as used in medical practice, beat morphology and image deformation. A binary classifier for each key cardiac pathology was trained using these features, extracted from the labeled ECGs from The Challenge. The binary classifiers were combined into a multi-label classifier by learning thresholds on the scores of the binary classifiers using Bayesian optimization in a cross-validation setting.

Results: our contribution submitted for evaluation achieved a challenge metric score of 0.28, 0.31, 0.32, 0.28 and 0.23 placing us (team DSC) 29, 25, 25, 28 and 28 out of 39 teams which submitted an official entry on 12-, 6-, 4-, 3- and 2-lead test datasets respectively.

1. Introduction

In The PhysioNet/Computing in Cardiology Challenge 2020 [1, 2] (hereafter The 2020 Challenge), participants were challenged to develop open-source algorithms to automatically identify cardiac abnormalities in 12-lead ECG recordings.

Increasing popularity of wearable health monitoring technology paves the way to the application of ECG analysis on a larger scale [3].

However, continuous monitoring of the 12-lead ECG is impractical and unattractive for a wearable system, due to the obtrusiveness and discomfort that the placement and connection of 10 electrodes would cause [4].

The PhysioNet/Computing in Cardiology Challenge 2021 [1, 5] (hereafter The Challenge or The 2021 Challenge) focused on automated, open-source approaches for classifying cardiac abnormalities from reduced-lead ECGs.

To this end, several databases are provided to the challenge participants: the CPSC2018 database [6], the CPSC2018-Extra database [2, 5], the INCART database [7], the PTB database [8], the PTB-XL database [9], the Chapman-Shaoxing Database [10], the Ningbo Database [11] and the Georgia database [2, 5]. For a description of these databases as well as for a description of the databases that were used for validation and testing, and therefore not disclosed, see [2, 5].

Our best entry in The Challenge uses features based on beat morphology and image deformation to train a binary one-versus-rest classifier for each cardiac abnormality.

2. Methods

For The 2021 Challenge we used five models described below.

The Chapman-Shaoxing and the Ningbo Database were excluded for training due to time limit requirements for training.

2020 model. Our contribution to The 2020 Challenge [12] uses features available online, supplemented by handcrafted features, which are specific to the pathologies provided in The 2020 Challenge. Some features are based on morphology, while others are based on significant ECG points. Using these features, binary classifiers on 2, 3, 4, 6 and 12 leads were trained for the scored pathologies and for frequently occurring combinations of pathologies. These classifiers were then combined in a hierarchical and parallel way. Thresholds on the scores output by the classifiers were chosen such that the distribution of predicted labels in holdout data of the training set was the same as the observed one in the training set.

Participation in The 2020 Challenge led to the observation that there is room for improvement in the robustness of the model, because there is a large variation in challenge metric score over different combinations of training- and test-sets. The challenge metric score was 0.616 on the 2020 validation set and 0.194 on the 2020 test set.

feature / class	AF	AFL	BBB	Brady	LBBB	RBBB	IRBBB	LAD	LAnFB	LPR	LQRSV	LQT	NSIVCB	NSR	PAC	PR	PVC	RAD	Stach	Tab	Tinv	Leads
Mean RR-interval length	x	x		x										x		x			x			I
Standard deviation of RR-interval length	x	x												x		x						I
Mean length of P waves														x	x		x					I
Distance between Q and S			x		x	x	x						x	x								I
Distance between start of P and Q	x													x								I
Morphology of the signal														x	x		x					A
Distance between start and peak of P and R	x									x				x								I, II
Distance between start of P and Q														x								I, II
Distance between end of P and R										x				x								I, II
Distance between end of T and beginning of next P														x								I, II
Distance between Q and S			x		x	x	x						x	x								I, II
Heart rate				x										x					x			I, II
Heart rate standard deviation	x	x		x										x		x			x			I, II
Difference between amplitude at Q and S								x	x					x				x				I
Distance between P and Q														x	x		x					I, II
Distance between Q and end of T												x		x								I, II
Inverted T (y/n)			x		x	x	x						x	x							x	A (I, II)
RMSE of a linear fit of the T-wave														x						x		A (I, II)
Difference between amplitude at Q and R								x	x					x				x				I-III, aVF
Amplitude of R compared to Q and S											x			x								aV*, V* (I, II)

Table 1: For each class, the features used to identify it and the leads from which the features are determined. 'A' indicates that all available leads are used. For the features related to the T-wave, leads I and II are used for the detection of the T-wave location. All leads are used to detect if the T-wave is inverted. This case is indicated in the table with (I, II). All features are represented by a single number, except the morphology of the signal, which is represented by 20 numbers.

The *2020 model* uses 24 general features and an additional 217 features per used lead.

Minimal model. We tried to improve the robustness of the *2020 model* by reducing the number of features and by reducing the number of binary classifiers.

We reduce the number of features by using, for a particular binary classifier, only those features which we manually a-priori deem as relevant for that particular binary classification problem (see Table 1).

The total number of features generated in the *minimal model* is 33, 36, 40, 49, and 73 for 2-, 3-, 4-, 6-, and 12-lead ECG's respectively, which include two general features (age and sex). The features for the *minimal model* are shown in the table.

We reduce the number of classifiers by only training one-versus-rest classifiers for the pathologies found in the reward matrix for The Challenge.

Minimal, bayesopt. As in [13], we maximize the challenge metric score by optimizing the thresholds on the binary-models' scores using the MATLAB® [14] implementation of Bayesian optimization [15].

Image deformation. This approach is based on scoring each beat, defined as the ECG signal within 3 consecutive R peaks.

First the signal is re-scaled to 250 Hz and possible trend in the signal baseline is removed using a median filter. Then for each available lead the R peaks are estimated using the 'gqrs' [16] program. The R peak location estimations from different leads are used to make a final R-location prediction. Each heartbeat is extracted and scaled to 60 heartbeats per minute (500 points). A simple estimator of beat morphology is implemented to find the possible location of the QRS interval, the T-wave and the P-wave.

A database of validated 'obvious' beats for several pathologies was extracted after manual review using a proprietary tool for ECG signal labeling (written in MATLAB® [14]). We call these validated beats atlases. We assume that each atlas is a truthful representation of a labeled pathology.

The one-dimensional image deformation distance to a random sample of 10 atlases per pathology is calculated for each new beat. To capture variation, the distances (Euclidean distance, cosine distance, Pearson's correlation distance) of each beat to the mean beat of the strip are computed. Distances are only computed for lead I, in order not to exceed the restrictions on running time.

To include information regarding the signal itself we compute the discrete one-dimensional wavelet transform of the data and collect the approximation coefficients. We use the Daubechies 1 wavelet [17]. The final training data set is comprised of 291 features: 4 based on heartbeats per minute, 6 based on distance of a beat to the mean beat of the strip, 180 based on image deformation, 37 based on beat morphology, and 64 based on the discrete wavelet transform.

To account for multi-pathology strips, we copy each strip with more than one label as two separate instances/strips with individual labels. We favored XGBoost [18] as a machine learning tool because its implementation allowed for easy supply of weights on each observation. We weigh each strip by the reciprocal of the amount of pathologies in this strip to preserve the prior distributions. In order to deal with unscored pathologies we combine them into a single pathology.

Once the XGBoost cross-validation (CV) procedure is finished for each pathology we take the mean predicted score for each beat across the strip. This produces our score on strip level. For each pathology we automatically select the score threshold based on the minimum of a simple cost function - 5*FN+1*FP. Each pathology is left with its own score threshold. The CV-AUC (Area Under the Curve) for the chosen thresholds for each pathology is shown in Table 6. No grid search for hyper parameters is run due to computational constraints. The key XGBoost hyper parameters used were: objective = multi:softprob, max depth = 8, estimators = 200.

3. Results

Table 2 shows the performance on the validation set for our (two) submissions during the official phase of The Challenge.

Table 3 shows the cross-validation challenge metric scores on the training data for all models described in Section 2. The *image deformation model* received the highest score (0.772). Only two image deformation models were built: one for 12 lead and one for 2 lead strips. The number of leads affects only the R peaks estimation and the beat extraction. The features for each beat are derived only from one lead (lead I).

Table 4 shows the challenge metric scores when we leave one dataset out for training and use that dataset for testing. As noted before, the Ningbo and Chapman-Shaoxing databases are only used for testing.

Table 5 shows the performance of our final selected entry (the *minimal model*).

Table 6 shows the AUC metric from cross-validation on the training set (CV-AUC).

model / # leads	12	6	4	3	2
2020 model	0.551	0.529	0.538	0.535	0.524
minimal model	0.385	0.375	0.378	0.383	0.365

Table 2: Challenge metric scores on validation set.

model / # leads	12	6	4	3	2
2020 model	0.532	0.501	0.506	0.502	0.496
2020, bayesopt	0.523	0.495	0.493	0.490	0.492
minimal model	0.315	0.309	0.313	0.315	0.315
minimal, bayesopt	0.375	0.366	0.364	0.368	0.359
image deformation*	0.772				0.758

Table 3: Challenge metric scores with 5-fold CV on training data excluding Chapman-Shaoxing and Ningbo.
*The CV dataset for image deformation consists of 3500 strips selected at random from the dataset with a constraint of minimum 80 examples of a pathology.

out-of-sample dataset / model	2020 model	minimal, bayesopt
CPSC	0.289	0.299
PTB	-0.356	-0.169
Georgia	0.246	0.241
Ningbo	0.408	0.335
Chapman-Shaoxing	0.542	0.428

Table 4: Challenge metric scores on leave-one-dataset-out evaluation with 12-lead ECG signals.

4. Discussion and Conclusions

Our main contributions are the *minimal model* and the *image deformation model*. The main idea behind the *minimal model* is to use only a relatively small set of handcrafted beat-morphology features. The 1-dimensional *image deformation model* relies additionally on features that are based on geometrical similarities between ECG strips. One big advantage with this approach is to use features that are easy to interpret and bear medical meaning. Further research can concentrate on filtering out ECG strips containing noise prior to feature-generation to achieve even greater robustness of the predictions.

Leads	Training	Validation	Test	Ranking
12	0.32 ± 0.01	0.39	0.28	29
6	0.31 ± 0.02	0.38	0.31	25
4	0.31 ± 0.01	0.38	0.32	25
3	0.32 ± 0.02	0.38	0.28	28
2	0.32 ± 0.02	0.37	0.23	28

Table 5: Challenge scores for our final selected entry (team DSC) using 5-fold cross-validation on the public training data excluding Chapman-Shaoxing and Ningbo, 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.

Class	2020 model	minimal, bayesopt	difference	image deform.*
AF	0.98	0.97	0.01	0.99
AFL	0.95	0.82	0.13	0.99
BBB	0.50	0.50	0.00	0.99
Brady	0.97	0.93	0.05	0.78
LBBB	0.98	0.98	0.00	1.00
RBBB	0.98	0.89	0.09	0.95
IAVB	0.94	0.93	0.01	0.98
IRBBB	0.92	0.76	0.16	0.95
LAD	0.94	0.87	0.07	0.94
LAnFB	0.97	0.91	0.06	0.97
LPR	0.95	0.88	0.07	0.98
LQRSV	0.92	0.83	0.10	0.98
LQT	0.94	0.84	0.10	0.98
NSIVCB	0.79	0.75	0.04	0.96
NSR	0.94	0.90	0.05	0.91
PAC	0.90	0.77	0.13	0.96
PR	1.00	0.80	0.19	0.98
PVC	0.84	0.66	0.18	0.97
Qab	0.78	0.61	0.17	0.96
RAD	0.97	0.89	0.08	0.99
SA	0.96	0.93	0.03	0.98
SB	0.99	0.96	0.03	0.99
Stach	0.99	0.99	0.01	1.00
Tab	0.87	0.73	0.13	0.96
Tinv	0.83	0.65	0.18	0.97

Table 6: CV-AUC on the 12-lead ECG's in training data. *The CV dataset for image deformation consists of 3500 strips selected at random from the dataset with a constraint of minimum 80 examples of a pathology.

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. Physiological Measurement 2020;41.

[3] Duncker D, Ding WY, Etheridge S, Noseworthy PA, Veltmann C, Yao X, Bunch TJ, Gupta D. Smart Wearables for Cardiac Monitoring—Real-World Use beyond Atrial Fibrillation. Sensors 2021;21(7).

[4] Figueiredo C, Mendes P. Towards Wearable and Continuous 12-Lead Electrocardiogram Monitoring: Synthesis of the 12-Lead Electrocardiogram Using 3 Wireless Single-Lead Sensors. Proceedings of the International Conference on Biomedical Electronics and Devices 01 2012;329–332.

[5] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, Rad BA, Elola A, Seyedi S, Ansari S, Li Q, Sharma A, Clifford GD. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[6] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, Liu Y, Ma C, Wei S, He Z, Li J, Kwee ENY. 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.

[7] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. PhysioBank PhysioToolkit and PhysioNet 2008;.

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

[9] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[10] 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.

[11] Zheng J, Cui H, Struppa D, Zhang J, Yacoub SM, El-Askary H, Chang A, Ehwerhemuepha L, Abudayyeh I, Barrett A, Fu G, Yao H, Li D, Guo H, Rakovski C. Optimal Multi-Stage Arrhythmia Classification Approach. Scientific Data 2020;10(2898):1–17.

[12] Nalbantov G, Ivanov S, Van Prehn J. Multi-Class Classification of Pathologies Found on Short ECG Signals. In 2020 Computing in Cardiology. 2020; 1–4.

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

[14] MATLAB. version 9.10 (R2021a). Natick, Massachusetts: The MathWorks Inc., 2021.

[15] Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. In Proceedings of the 25th International Conference on Neural Information Processing Systems, volume 2 of NIPS'12. 2012; 2951–2959.

[16] Silva I, Moody G. An Open-source Toolbox for Analysing and Processing PhysioNet Databases in MATLAB and Octave. Journal of Open Research Software 09 2014;2:e27.

[17] Daubechies I. Ten Lectures on Wavelets. SIAM, 1992.

[18] 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. 2016; 785–794.

Address for correspondence:

Jeffrey van Prehn
Lippestraat 24, Voerendaal, The Netherlands
jvanprehn@gmail.com