Classification of 12-lead ECGs Using Gradient Boosting on Features Acquired With Domain-Specific and Domain-Agnostic Methods

Durmus Umutcan Uguz[1], Felix Berief[1], Steffen Leonhardt[1], Christoph Hoog Antink[1]

[1]Medical Information Technology, Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Aachen, Germany

Abstract

This year, the objective of the PhysioNet/Computing in Cardiology challenge was the classification of 12-lead electrocardiograms (ECG). The approach presented in this paper consists of two parts, feature extraction and classification. The extracted features can be separated into domain-specific and domain-agnostic features, where domain-specific features are based on known ECG processing methods such as QRS-detectors. Domain-agnostic features are generated by wavelet transforms that take the raw 12-lead ECG as input. Additionally, a novel beat-to-beat correlation analysis is proposed to identify arrhythmia occurring among other healthy beats. These features are then combined and classified by gradient-boosted trees implemented in Python. To account for the complexity of the multi-label and multi-class problem definition, a One-vs-Rest scheme is utilized, where distinct classifiers for each class determine whether a sample belongs to said class. The resulting imbalance in training sets for each classifier was compensated for by giving the positive samples a higher weight. The classifiers were trained using the XGBoost gradient boosting system. The proposed classification scheme of the team "desafinado" received a score of 0.576 on the validation dataset and a score of 0.233 on the test set of the challenge (rank 19 of 41).

1. Introduction

Cardiovascular diseases are the leading causes of mortality and account for $48\,\%$ of all deaths among the non-communicable diseases [1]. Advances in the electrocardiogram (ECG) monitoring in recent years have increased the demand for automated and computer-aided ECG diagnosis. Machine learning tools are widely employed to meet this demand, have been tested in the problem of arrhythmia classification [2,3] and remain to be a focal point in the future of ECG diagnosis. The necessity of an automated arrhythmia classification is addressed in the PhysioNet/Computing in Cardiology Challenge 2020 [4] to encourage open-source approaches and obtain reproducible results.

This paper presents the approach developed by our team "desafinado", which uses gradient boosting on features generated from fiducial points, wavelet transform, higher order statistics and a novel beat-to-beat correlation analysis of the 12-lead ECG data.

2. Methods

In the following, preprocessing, feature extraction, classification, and postprocessing are described.

2.1. Preprocessing

ECG recordings were low-pass filtered ($80\,\mathrm{Hz}$), after removing the baseline wander by using a two-step median filter (lengths of $200\,\mathrm{ms}$ and $600\,\mathrm{ms}$). To reduce the complexity of the feature vector from the beat-to-beat correlation analysis in Section 2.2.3, Kors transformation was used to transform 12-lead ECG recordings into orthogonal XYZ leads [5]. Both 12-lead ECG and obtained XYZ leads were segmented using BioSPPy [6].

2.2. Feature Extraction

For the classification, a wide range of features was used. These can be categorized into three groups.

2.2.1. ECG Timing Features

After segmenting the beats using Pan&Tompkins algorithm [7], following features ($\vec{v}_{\mathrm{timing}}$) were calculated from the resulting R-peak locations r_i:

- RR-interval $\delta_i = r_{i+1} - r_i$,
- ΔRR $\delta_i^2 = \delta_{i+1} - \delta_i$,
- pNN50,
- pNN20,
- number of beats.

In addition to these features, the following timing features were calculated from the segmented heartbeats:

 ISSN: 2325-887X DOI: 10.22489/CinC.2020.363

- P-wave onset,
- QT interval,
- ST interval,
- T-wave onset,
- QRS width.

2.2.2. ECG Waveform Features

From the segmented heartbeats, the amplitudes of R, P, Q, T, and S-wave were calculated. In addition to these features calculated for 12 leads, the ratio of the amplitudes from the following waves were included:

- R/P,
- R/Q,
- R/T,
- R/S.

The set of waveform features ($\vec{v}_{\text{wav}}$) is completed by adding the average signal quality [8], average energy in the signal and the approximation coefficients of 3$^{\text{rd}}$-level wavelet decomposition using a $db1$ wavelet.

2.2.3. Beat-to-Beat Correlation Analysis

Using an analysis window of $600\,\text{ms}$ centered at the R-peak locations r_i, the median beat was calculated for each XYZ lead. Using this segmentation, beat waveforms were extracted for an analysis window of $600\,\text{ms}$ centered at r_i. Using these segmented heartbeats, the median beat for each lead was calculated. Using Pearson's correlation, a correlation vector consisting of ρ_i elements was calculated as

$$\rho_i = \frac{1}{L-1}\sum_{n=1}^{L}\Big(\frac{x_i[n]-\mu_i}{\sigma_i}\Big)\Big(\frac{x_{\text{med}}[n]-\mu_{\text{med}}}{\sigma_{\text{med}}}\Big), \quad (1)$$

where ρ_i, μ_i and σ_i represent the correlation coefficient, the mean and the standard deviation of the i^{th} beat, respectively. $x_i[n]$ is the i^{th} beat among N beats from the segmentation and $x_{\text{med}}[n]$ is the median beat of these N beats.

Segmented heartbeats were then sorted according to their ρ_i with the median heartbeat of the section. The most correlated (*accordant*) heartbeat and the least correlated (*dissonant*) heartbeat were retrieved, as illustrated in Figure 1. This attempt aims to isolate arrhythmias that do not occur at each heart cycle, and are thus prominent with their dissimilarity to the neighboring beats.

The following features were calculated from the correlation vector $\vec{\rho}$ for XYZ lead:

- std, mean, median, max, min of $\vec{\rho}$,
- std($\vec{\rho}$)/mean($\vec{\rho}$),
- std($\Delta\rho$), where $\Delta\rho = \rho_{i+1} - \rho_i$,
- std($\Delta\rho$)/mean($\vec{\rho}$).

In addition to these features from the beat-to-beat correlation vector $\vec{\rho}$, the following waveform features were calculated for both the dissonant and the accordant beats at each XYZ lead:

- skewness and kurtosis of each $50\,\text{ms}$ long section,
- argmax, max and min for the region [$0\,\text{ms}$ to $110\,\text{ms}$],
- max and min for the region [$240\,\text{ms}$ to $360\,\text{ms}$],
- argmax, max and min for the region [$410\,\text{ms}$ to $570\,\text{ms}$].

2.3. Classifier

The problem of assigning one or more of K classes to the 12-channel ECG recording and its corresponding features is a so-called multi-class multi-label task. The presented classifier first transforms this problem into K binary classification problems by using a One-vs-Rest approach, where the j^{th} classifier decides whether the ECG recording belongs to class j or not. This problem transformation was needed since the utilized gradient boosting classifier cannot handle multiple labels per input. Due to the application of the One-vs-Rest approach, binary classifiers receive highly imbalanced datasets even with slightly more positive samples [9]. To account for this imbalance, positive samples for class k were re-weighted during the training of classifier k by

$$w_k = \frac{N_{\text{tot}} - N_k}{N_k}, \quad (2)$$

where N_k and N_{tot} are the number of samples for class k and the total number of samples, respectively. The negative samples received a weight of 1. Classes considered to be equivalent by the organizers were treated as the same class and only the scored classes were used in the classifier. Thus, the total number of classes was $K = 25$.

The classification was handled by the XGBoost algorithm utilizing the idea of gradient boosted trees [10]. Boosted trees are decision trees used as an ensemble to build a single stronger classifier from many weak classifiers. Gradient boosting refers to the gradient descent method employed to find the best decision tree, whereas boosting itself indicates that samples previously misclassified receive larger weights for the next training steps. XGBoost utilize a variety of highly optimized techniques to make the training both robust and fast while still maintaining a high degree of customizability.

Among 20 tunable parameters offered by XGBoost, 3 of the most impactful ones were tested extensively for this challenge. All of these parameters, namely *gamma*, *min_child_weight* and *max_depth*, play an important role for avoiding overfitting without losing the ability to generalize to unseen data. As the hidden test data partly includes samples from a new data source, the parameter combinations were optimized not only through 5-fold cross-validation but also by holding out entire data sources for evaluation.

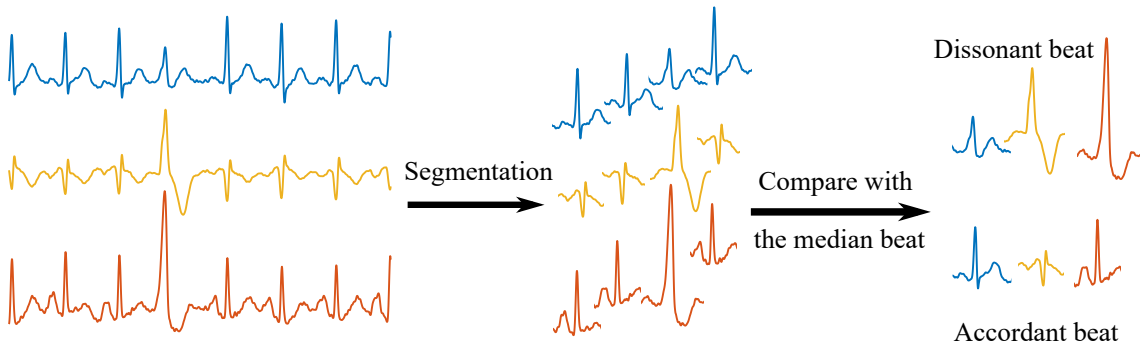

Figure 1. The correlation analysis conducted on the orthogonal XYZ leads. After segmenting the heartbeats, the found median beat is Pearson-correlated with each individual beat to find the least and the most similar beats, called dissonant and accordant beats, respectively.

2.4. Postprocessing

Since the classifier makes use of the One-vs-Rest technique, it consists of independent binary classifiers for each of the K classes. One of the downsides of this approach is the independent decisions as some diagnoses have a higher probability to occur jointly whereas other diagnoses might not be able to appear together in the same recording. This could also lead to a scenario where no output class is assigned.

In order to reduce the impact of these limitations, each class was assigned a specific threshold of t_k instead of relying on the default binary classification threshold of 0.5. The thresholds t_k were found by optimizing the challenge score as a function of t_k during cross-validation.

3. Results

Table 1 shows the results of the 10-fold cross-validation using the training data (top part) as well as the results of the challenge validation (bottom part), i.e. the result of the official phase of the challenge. Results are presented in terms of area under the receiver operating characteristic (AUROC), area under the precision-recall curve (AUPRC), accuracy, F-measure, and the challenge score as described in [4]. Moreover, mean and standard deviation (SD) over all folds are calculated.

The cross-validation shows minimal variability between the different folds. Moreover, while the AUROC and AUPRC are higher on the challenge validation dataset, accuracy and F-measure are lower compared to the average results of the cross-validation. The challenge validation score, however, lies within the standard deviation of the cross-validation results. On the evaluation system of the challenge, the runtime was 101 hours and 28 minutes. During the official phase of the challenge, the algorithm ranked 95 out of 306.

10-fold Training Cross-Validation

indx	AUROC	AUPRC	Accuracy	F-measure	Score
1	0.83	0.30	0.29	0.45	0.57
2	0.84	0.30	0.30	0.46	0.58
3	0.83	0.31	0.30	0.46	0.58
4	0.83	0.29	0.28	0.45	0.57
5	0.83	0.30	0.29	0.45	0.57
6	0.83	0.29	0.30	0.45	0.57
7	0.84	0.30	0.29	0.45	0.57
8	0.83	0.28	0.28	0.44	0.55
9	0.83	0.29	0.28	0.45	0.56
10	0.83	0.29	0.30	0.45	0.57
mean	0.83	0.30	0.29	0.45	0.57
sd	0.0030	0.0064	0.0074	0.0069	0.0088
Challenge Validation					
	0.906	0.478	0.224	0.413	0.576

Table 1. Results of 10-fold cross-validation combining all training datasets (top) as well as challenge validation results (bottom).

Table 2 shows the results of the leave-one-dataset-out cross-validation using the training data, Table 3 shows the results achieved on the final challenge test set.

Dataset	AUROC	AUPRC	Accuracy	F-measure	Score
CPSC unused	0.63	0.10	0.06	0.14	0.41
CPSC	0.88	0.54	0.16	0.17	0.48
Georgia	0.72	0.22	0.14	0.29	0.34
PTB	0.59	0.08	0.00	0.02	-3.26
PTB-XL	0.74	0.23	0.25	0.30	0.03
StPetersburg	0.54	0.07	0.00	0.03	0.20
mean	0.68	0.21	0.10	0.16	-0.30
sd	0.1208	0.1787	0.0981	0.1218	1.4584
mean \PTB	0.70	0.23	0.12	0.19	0.29
sd \PTB	0.1258	0.1875	0.0950	0.1125	0.1762

Table 2. Results of leave-one-dataset-out cross-validation on the training data.

Dataset	AUROC	AUPRC	Accuracy	F-measure	Score
Test Database 1	0.967	0.833	0.325	0.182	0.681
Test Database 2	0.887	0.481	0.186	0.412	0.556
Test Database 3	0.806	0.358	0.009	0.202	-0.013
Full Test Set	0.822	0.362	0.089	0.298	0.233

Table 3. Results of the official, final challenge test evaluation.

First, it becomes obvious that the performance degrades when complete datasets are held out for testing. Moreover, great variability between datasets is observed. In particular, the results on the "PTB" dataset are most inferior. However, even when this dataset is left out, the average score degrades to 0.29 (as compared to 0.57 in both the 10-fold training cross-validation and the challenge validation).

Similar observations can be made in terms of the final, official challenge test evaluation (Table 3): While the scores achieved on the hidden test databases 1 and 2 are comparatively high (0.681 and 0.556), the score on database 3 is low (-0.013). On the full test set, the submitted algorithm of our team "desafinado" received an overall score of 0.233, ranking it 19th out of 41 final entries.

4. Discussion and Outlook

The presented approach demonstrates that by using straightforward feature engineering and sophisticated ensemble machine learning, a competitive algorithm for classification of 12-lead ECGs can be constructed. Both in terms of rank (19 out of 41) and in terms of score (0.233 vs. a numeric average of 0.174 over all 41 teams), it achieved above-average results in the challenge.

Notwithstanding, it is obvious that several approaches achieved significantly better results, the top four teams achieving scores more than twice as high. In addition, it can be concluded that the ability to generalize is limited, as obvious by the variation in results on the test databases.

In future work, the proposed features need to be optimized. It is obvious that large redundancy is currently not exploited properly. For example, the RR-intervals are calculated for each lead separately instead of using a fused approach. Finally, the used classifier allows to analyze the feature importance. This property will be exploited in future work to allow further analysis on how the classifications are achieved. It may also help to shed light on which combination of leads and features are most important and may be optimized for improved results.

References

[1] Mendis S, Puska P, Norrving B. Global Atlas on Cardiovascular Disease Prevention and Control. World Health Organization, 2011.

[2] Sahoo S, Dash M, Behera S, Sabut S. Machine Learning Approach to Detect Cardiac Arrhythmias in ECG Signals: A Survey. IRBM 2020;41(4).

[3] Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A Review on Deep Learning Methods for ECG Arrhythmia Classification. Expert Systems with Applications X 2020; 7:100033.

[4] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad AB, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol Meas 2020;In Press.

[5] Kors JA, van Herpen G, Sittig AC, van Bemmel JH. Reconstruction of the Frank Vectorcardiogram from Standard Electrocardiographic Leads: Diagnostic Comparison of Different Methods. Eur Heart J 1990;11(12):1083–1092.

[6] Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, Fred A, et al. BioSPPy: Biosignal processing in Python, 2015–. URL `https://github.com/PIA-Group/BioSPPy/`. [Online; accessed 2020-09-30].

[7] Pan J, Tompkins WJ. A real-time QRS Detection Algorithm. IEEE Trans Biomed Eng 1985;(3):230–236.

[8] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, S H Chen A. NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing, 2020. URL `https://github.com/neuropsychology/NeuroKit`.

[9] Luque A, Carrasco A, Martín A, de las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition 2019;91:216–231.

[10] 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. 2016; 785–794.

Address for correspondence:

Durmus Umutcan Uguz
Medical Information Technology
RWTH Aachen University
Pauwelsstr. 20, 52074 Aachen, Germany
uguz@hia.rwth-aachen.de