Automatic Classification of 12-, 6-, 4-, 3-, and 2-Lead Electrocardiograms Using Morphological Feature Extraction

Alexander Hammer, Matthieu Scherpf, Hannes Ernst, Jonas Weiß, Daniel Schwensow, Martin Schmidt

Institute of Biomedical Engineering, TU Dresden, Dresden, Germany

Abstract

Cardiovascular diseases are the global leading cause of death. Automated electrocardiogram (ECG) analysis can support clinicians to identify abnormal excitation of the heart and prevent premature cardiovascular death. An explainable classification is particularly important for support systems. Our contribution to the PhysioNet/CinC Challenge 2021 (team name: ibmtPeakyFinders) therefore pursues an approach that is based on interpretable features to be as explainable as possible.

To meet the challenge goal of developing an algorithm that works for both 12-lead and reduced lead ECGs, we processed each lead separately. We focused on signal processing techniques based on template delineation that yield the template's fiducial points to take the ECG waveform morphology into account. In addition to beat intervals and amplitudes obtained from the template, various heart rate variability and QT interval variability features were extracted and supplemented by signal quality indices. Our classification approach utilized a decision tree ensemble in a one-vs-rest approach. The model parameters were determined using an extensive grid search.

Our approach achieved challenge scores of 0.47, 0.47, 0.40, 0.43, and 0.45 on 12-, 6-, 3-, 4-, and 2-lead validation sets, respectively.

1. Introduction

Cardiovascular diseases are the global leading cause of death [1]. The analysis of electrocardiograms (ECGs) can be used to non-invasively detect anomalies in the electrical impulse formation and conduction of the heart as predictors for cardiovascular diseases [2]. The 12-lead ECG is considered the clinical standard. Devices with fewer leads are cheaper and easier to use but provide less information. Deep learning approaches have already demonstrated great potential in automated 12- and reduced lead ECG analysis and anomaly detection [3]. However, deep learning approaches are often criticized for their lack of comprehensibility regarding decision-making (so-called black box) [4]. Since physicians are responsible for their diagnosis, explainable classification based on clinically interpretable features is required instead of inappropriate black box characteristics [4]. Therefore in this study, as part of the PhysioNet/Computing in Cardiology (CinC) 2021 challenge [5], which focused on automated, open-source approaches for classifying cardiac anomalies from reduced- or full-lead ECGs, we demonstrate the potential of manual feature extraction in combination with a boosted decision tree ensemble for this purpose.

2. Methods

To achieve the most comprehensible classification of ECGs with different numbers of leads, we have extracted various clinically interpretable features, including heart rate variability (HRV), morphological features from delineated template beats, beat-to-beat QT interval variability (QTV), and signal quality indices (SQIs) from each lead. A boosted decision-tree ensemble (DTE) was trained with these features, supplemented by global features. The process of training and classification is shown in Figure 1. The entire implementation was done with Matlab 2020b (MathWorks Inc., Natick, Massachusetts, USA).

2.1. ECG Preprocessing

A band-pass filter (passband: 0.3 to 35 Hz) was applied to all ECGs. To avoid boundary effects caused by filtering, zero-padding was added. SQIs were extracted from filtered ECGs. For other features, QRS complexes were detected subsequently using an algorithm which is based on Hilbert transformation and included in the *biosig* toolbox [6].

We corrected erronously detected QRS complexes to make our approach more robust against artifacts and anomalies in ECGs. This included discarding peaks within the first or last 300 ms, and shifting detected QRS complexes to local maxima within ± 50 ms windows around the QRS complexes. Further, sporadic positive/negative QRS complexes in the majority of QRS complexes with

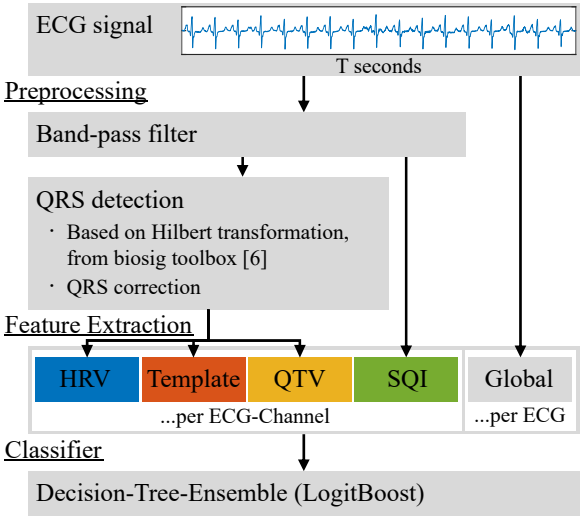

Figure 1. Processing pipeline of the ECG classification.

the opposite sign were rejected as well as QRS complexes with an absolute amplitude height lower than 100 mV or greater than half of the maximum peak height if they occurred within a frequency of less than 0.2 Hz. The occurrence of the latter was included as a feature. Finally, the QRS complex with greater deviation from the mean QRS amplitude level was rejected if two QRS complexes occurred with a distance of less than 250 ms.

2.2. Feature Extraction

In addition to global features (age, sex, signal length, and gain), we extracted 68 HRV-, template, and QTV-features, as well as SQIs, from each ECG lead separately, taking the entire signal into account.

2.2.1. HRV-Feature Extraction

RR intervals were extracted from the distances between QRS complexes and filtered for physiological and non-physiological RR intervals by using the filtRR function from *PhysioZoo* toolbox [7]. The filtration rate (filtRate) was included in the feature set. The physiological RR intervals were used for the calculation of statistical, geometric, non-linear, and frequency-based HRV features described in [8]. Mean, median, minimum, and maximum RR intervals completed the HRV feature set.

2.2.2. Template-Feature Extraction

To take the ECG waveform morphology into account, we focused on signal processing techniques based on template delineation that yield the template's fiducial points. Therefore, based on the formula from Laguna et al. [9], windows for beat extraction were defined from the start of the P wave (Pon) to the end of the T wave (Tend) as

$$t_{Pon,j} = t_{QRS,j} - 370\,\text{ms} \tag{1}$$

$$t_{Tend,j} = \begin{cases} t_{QRS,j+1} - 240\,\text{ms}, & \overline{RR} \geq 720\,\text{ms} \\ t_{QRS,j} + 2/3\overline{RR}, & \overline{RR} < 720\,\text{ms} \end{cases}, \tag{2}$$

where $t_{Pon,j}$ is the begin and $t_{Tend,j}$ the end of the PT interval, $\overline{RR}$ is the average RR interval and $t_{QRS,j}$ is the location of the QRS complex of beat j, assuming a maximum QRS- and PQ interval length of 370 ms [10].

Two clusters were formed from the beats to separate normal from abnormal beats. The clustered beats were shifted against each other up to the maximum cross-correlation and then averaged to generate the templates. Subsequently, characteristic fiducial points were determined based on Laguna et al. [9] for both templates to calculate features in the form of intervals between fiducial points or amplitude heights (see Figure 2). In addition, the mean difference between the two templates (clDiff) and the number of beats per cluster (clSize1/2) were used as features.

2.2.3. QTV-Feature Extraction

Two-dimensional signal warping (2DSW) was used for QTV-feature extraction [11]. The 2DSW algorithm achieves a robust estimate of time intervals on a beat-by-beat basis. For this purpose, a 2D grid of warping points is placed over the (regular) template. By moving the folding points in x and y direction, the template is adapted to each beat separately while minimizing the Euclidean distance between the template and the beat. Hence, beat-to-beat changes in the annotated features can be tracked.

We used standard QTV features including standard deviation of QT intervals and the QT interval variability index [12]. To consider the inverse relationship between QTV and T wave amplitude, we included the T wave amplitude-corrected measures cQTV, cSDQT, and cQTVi in the feature set [13]. The QTV feature set was supplemented by the mean, median, minimum, and maximum QT length (meanQT, medianQT, minQT, maxQT).

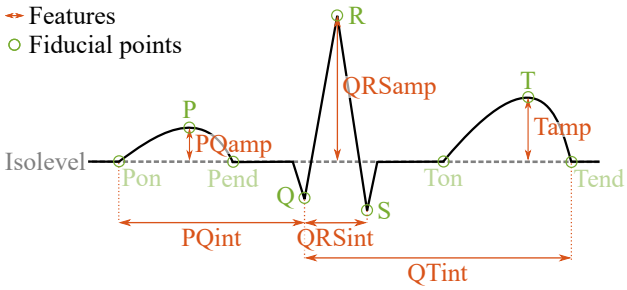

Figure 2. Fiducial points and extracted Template-features.

2.2.4. SQI-Feature Extraction

To integrate the dimension of signal quality into our approach for automated ECG analysis, we calculated various SQIs that are pooled in the *fecgsyn* toolbox [14].

2.3. Classification

The task of the PhysioNet/CinC Challenge 2021 was to provide an algorithm that works for detecting 30 anomalies in full 12-lead ECGs as well as in settings of 2- (I, II), 3- (I, II, V2), 4- (I-III, V2), and 6-lead (I-III, aVR, aVL, aVF) ECGs. Accordingly, we trained a model for each setting with the extracted features of the corresponding leads.

Since a single ECG signal could contain several or none of the anomalies to be classified, we applied a one-vs-rest approach which consisted of a DTE trained with the adaptive logistic regression algorithm (LogitBoost) for each anomaly. A grid search was carried out to determine the optimal model parameters based on the public training set. We varied the number of learners between 128 and 768 and the number of splits per level from 2 to 4. The learning rate was set to 0.1 and the number of bins to 256. The models were evaluated using the F-measure, the area under the precision-recall curve (auprc), and the challenge metric.

To train each model, the public dataset was divided into stratified training and test sets in a ratio of 90 to 10. Due to the challenge memory limit of 100 GB, we had to forego cross-validation (cv), which tends to generate more robust models. The F-measure was used for classification threshold adaption.

More than 95% of the EKGs of the training set were 10 seconds at maximum, and over 99% were 144 seconds at maximum. With a length of 30 minutes each, 74 ECGs were significantly longer and were therefore not taken into account during training. Also, we merged anomaly classes, that were scored together, for training.

3. Results

According to the grid search (see Figure 3), the classification performance for both 12 and 2 leads generally increases with model complexity. The results of the most and less complex 12- and 2-lead models are representative for all settings. Due to training time and memory limitations within the challenge, we decided on 4 splits and 256 learners as a trade-off between complexity and classification performance.

On the public training set, we achieved a challenge score between 0.58 (2 leads) and 0.61 (12 leads) with stratified hold-out validation (see Table 1). With a stratified 3-fold cv, we achieved a slightly higher challenge score of 0.60 for 2 leads and 0.63 for 12 leads. Our challenge score on the hidden validation set reached from 0.40 (4 leads) to 0.47 (12 leads).

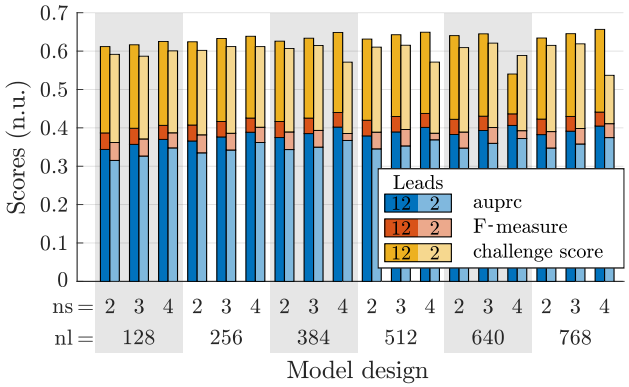

Figure 3. Results of grid search for models with different model design regarding the number of splits (ns) and number of learner (nl), based on public training set.

Generally, there is significant overlap (8 out of 10) between the most important features for 2- and 12-lead models, even if the leads, from which the features are calculated, differ. (see Figure 4). Leads I and II, which are included in all reduced lead sets, play a subordinate role in the 12-lead model.

Leads	Training		Validation	Test	Ranking
	3-fold cv	hold-out			
12	0.63	0.61	0.47	???	???
6	0.61	0.59	0.47	???	???
4	0.61	0.58	0.40	???	???
3	0.60	0.59	0.43	???	???
2	0.60	0.58	0.45	???	???

Table 1. Challenge scores for our final selected entry (team ibmtPeakyFinders) using 3-fold cross-validation (cv) or hold-out validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring as well as ranking on the hidden test set.

4. Discussion and Conclusions

Remarkably, the classification results with a reduced lead number are competitive with the results with 12 ECG leads, which indicates a great potential of mobile ECG patches such as those used in the TIMELY project[1]. However, it should be taken into account that the extracted features and methods are primarily developed for limb leads according to Einthoven (I-III), which are represented in all challenge lead settings. This contradicts the fact that many of the most important features of the 12-lead model were extracted from chest wall leads.

In this study, we were able to show that decent results

[1] http://www.timely-project.com/

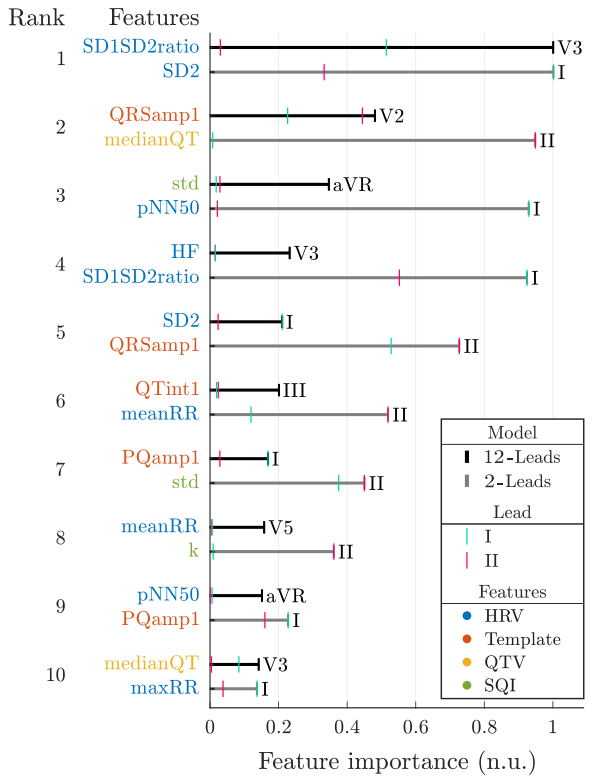

Figure 4. Ranking of the 10 most important features, including the leads of which they are calculated from, in 12- vs 2-lead model. The importance of the respective features, calculated from lead I and II, is marked separately.

can be achieved with a manual feature-based classification approach that is highly explainable. Due to a large discrepancy between the validation results and the significantly higher results in training, we see great potential in our approach with a more robust design of the classifier using cv-based training in future studies.

Acknowledgments

This study was partly supported by grants from the European Union's Horizon 2020 research and innovation programme (TIMELY, No 101017424) and the European Regional Development Fund (ERDF, No 100278533).

References

[1] World Health Organization. Cardiovascular diseases. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] 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):154. ISSN 2052-4463.

[3] Strodthoff N, Wagner P, Schaeffter T, Samek W. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. IEEE Journal of Biomedical and Health Informatics 2020;25(5):1519–1528.

[4] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery 2019;9(4):e1312. ISSN 1942-4795.

[5] Reyna M, Sadr N, Gu A, Perez Alday E, Liu C, Seyedi S, et al. Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021 (version 1.02). In Computing in Cardiology 2021, volume 48. 2021; 1–4.

[6] Vidaurre C, Sander TH, Schlögl A. BioSig: The free and open source software library for biomedical signal processing. Computational Intelligence and Neuroscience 2011; 2011:935364. ISSN 1687-5273.

[7] Behar JA, Rosenberg AA, Weiser-Bitoun I, Shemla O, Alexandrovich A, Konyukhov E, et al. PhysioZoo: A Novel Open Access Platform for Heart Rate Variability Analysis of Mammalian Electrocardiographic Data. Frontiers in Physiology 2018;9:1390. ISSN 1664-042X.

[8] Vollmer M. A robust, simple and reliable measure of heart rate variability using relative RR intervals. In 2015 Computing in Cardiology Conference (CinC). 2015; 609–612.

[9] Laguna P, Moody GB, García J, Goldberger AL, Mark RG. Analysis of the ST-T complex of the electrocardiogram using the Karhunen—Loeve transform: Adaptive monitoring and alternans detection. Medical Biological Engineering Computing 1999;37(2):175–189. ISSN 1741-0444.

[10] Schmidt M, Baumert M, Malberg H, Zaunseder S. Iterative two-dimensional signal warping—Towards a generalized approach for adaption of one-dimensional signals. Biomedical Signal Processing and Control 2018;43:311–319. ISSN 1746-8094.

[11] Schmidt M, Baumert M, Porta A, Malberg H, Zaunseder S. Two-Dimensional Warping for One-Dimensional Signals—Conceptual Framework and Application to ECG Processing. IEEE Transactions on Signal Processing 2014; 62(21):5577–5588. ISSN 1941-0476.

[12] Berger RD, Kasper EK, Baughman KL, Marban E, Calkins H, Tomaselli GF. Beat-to-Beat QT Interval Variability. Circulation 1997;96(5):1557–1565.

[13] Schmidt M, Baumert M, Malberg H, Zaunseder S. T Wave Amplitude Correction of QT Interval Variability for Improved Repolarization Lability Measurement. Frontiers in Physiology 2016;7:216. ISSN 1664-042X.

[14] Andreotti F, Behar J, Zaunseder S, Oster J, Clifford GD. An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms. Physiological Measurement 2016;37(5):627–648. ISSN 0967-3334.

Address for correspondence:

Alexander Hammer
Institute of Biomedical Engineering, TU Dresden
Fetscherstr. 29, Dresden, Germany
alexander.hammer@tu-dresden.de