Cardiac Pathologies Detection and Classification in 12-lead ECG

Radovan Smisek[1,2], Andrea Nemcova[1], Lucie Marsanova[1], Lukas Smital[1], Martin Vitek[1], Jiri Kozumplik[1]

[1]Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering, Brno, Czech Republic
[2]The Czech Academy of Sciences, Institute of Scientific Instruments, Brno, Czech Republic

Abstract

Background: Automatic detection and classification of cardiac abnormalities in ECG is one of the basic and often solved problems. The aim of this paper is to present a proposed algorithm for ECG classification into 19 classes. This algorithm was created within PhysioNet/CinC Challenge 2020, name of our team was HITTING.

Methods: Our algorithm detects each pathology separately according to the extracted features and created rules. Signals from the 6 databases were used. Detector of QRS complexes, T-waves and P-waves including detection of their boundaries was designed. Then, the most common morphology of the QRS was found in each record. All these QRS were averaged. Features were extracted from the averaged QRS and from intervals between detected points. Appropriate features and rules were set using classification trees.

Results: Our approach achieved a challenge validation score of 0.435, and full test score of 0.354, placing us 11 out of 41 in the official ranking.

Conclusion: The advantage of our algorithm is easy interpretation. It is obvious according to which features algorithm decided and what thresholds were set.

1. Introduction

This paper describes a fully automated algorithm for detection of cardiac pathologies. This algorithm was created within PhysioNet/CinC Challenge 2020. Detailed information about this competition including a description of the metrics for algorithm evaluation is in [1].

This algorithm is able to detect rhythm pathologies - atrial fibrillation (AF), atrial flutter (AFL), pacing rhythm (PR), sinus arrhythmia (SA), sinus tachycardia (STach) and sinus bradycardia (SB); conduction disturbances - 1st degree AV block (IAVB), prolonged PR interval (LPR), complete and incomplete right bundle branch block (RBBB and IRBBB), left bundle branch block (LBBB), left anterior fascicular block (LAnFB) and nonspecific intraventricular conduction disorder (NSIVCB); premature atrial and ventricular contraction (PAC and PVC); left and right axis deviation (LAD and RAD); low QRS (LQRSV) and T wave abnormal (TAb).

The success of the developed algorithms was determined on publicly available PhysioNet/CinC Challenge 2020 databases and on a hidden test set included in the PhysioNet/CinC Challenge 2020 [1]. This database contains 6 subdatabases - China Physiological Signal Challenge 2018 database (CPSC) [2], Southeast University in China database (CPSC 2), St Petersburg INCART 12-lead Arrhythmia Database [3], PTB Diagnostic ECG Database [3, 4], PTB-XL database [3, 5, 6] and Georgia database (Georgia). Bradycardia (Brady), prolonged QT interval (LQT), Q wave abnormal (QAb) and T wave inversion (TInv) that were scored in the challenge are not detected by our algorithm.

The name of our team in the challenge was HITTING.

2. Methods

The block diagram of the proposed algorithm is shown in the Figure 1. The first step of the algorithm is QRS detection. Detection of the P-wave and the T-wave and the extraction of features describing the QRS morphology are performed on the averaged PQRST segment. We average only those segments that have a dominant QRS morphology in the record. Segments of another QRS morphology are excluded because they are either pathological complexes or artifacts. Features for classification are extracted from the averaged QRS and from intervals between detected points (P, QRS and T and their onsets and offsets). Appropriate features and rules were set using classification trees.

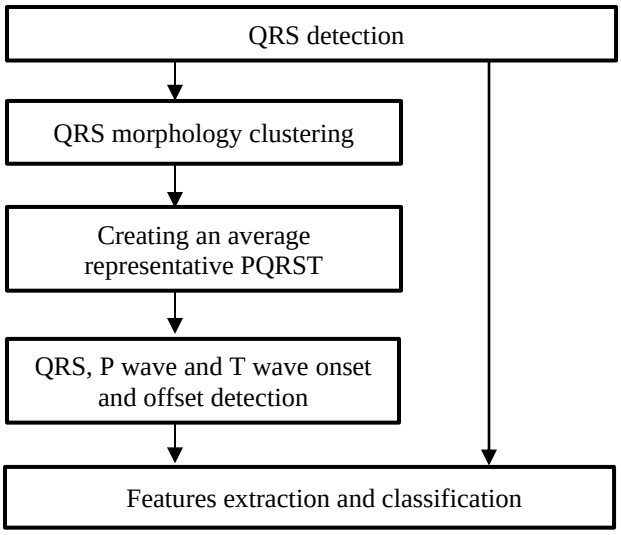

Figure 1. Block diagram of the ECG classification.

2.1. QRS detection

To obtain reliable QRS positions, the detection was performed using combination of 3 detectors – based on phasor transform, continuous wavelet transform, and S-transform. This detector is described in detail in [7].

2.2. QRS morphology clustering and averaging

The first step of the clustering is the alignment of QRS complexes. The cross-correlation between the individual QRS complexes is then calculated. The complexes are thus divided into clusters according to morphology. For detection of most pathologies, only the cluster containing QRS complexes of the most common morphology is preserved. The PQRST segments containing the QRS of the most frequent morphology are finally averaged.

2.3. QRS, P waves and T waves onsets and offsets detection

Detection of QRS, P waves and T waves onsets and offsets was performed using the averaged PQRST segment. For QRS onset and offset detection we used an algorithm based on wavelet transform and thresholding, specifically we used the bior 1.5 wavelet. The detector of QRS onsets and offsets was previously used in the challenge with name ISCE 2018 LBBB Initiative, where it was evaluated as the most accurate of all competing algorithms. This algorithm is described in [8]. We used an algorithm based on phasor transform to detect P waves including P waves boundaries. This algorithm is described in [9]. T waves were detected using wavelet transform and thresholding. Next, the tangent was simulated at the point of the steepest slope of the curve after the peak T. The offset of the T wave was determined at the point where this tangent intersects the zero isoline. T wave onsets were not detected because this point is not significant for any pathology.

2.3. Features extraction and classification

We extracted features potentially suitable for detection of any pathology. The most suitable features and thresholds were determined for some pathologies based on standard medical definitions of those pathologies (STach, SB, IAVB, LPR, LBBB, PR, NSIVCB, LQRSV and SNR). For other pathologies the most suitable features and thresholds were selected using classification trees (AF, AFL, SA, RBBB, IRBBB, LAnFB, LAD, RAD, PAC, PVC, and Tab).

Pathologies detected according to medical definitions were detected as follows:

- **IAVB and LPR** are the same pathology, so we merged them. IAVB and LPR were detected when PQ interval is longer than or equal to 205 ms. It is detected only when AF is not detected in the same record.
- **STach** was detected when heart rate is higher than 100 bpm. STach is detected only when AF is not detected in the same record.
- **SB:** was detected when heart rate is lower than 60 bpm. SB is detected only when AF is not detected in the same record.
- **LBBB** was detected according to the modified Strauss criteria. The criteria for LBBB include the presence of QS- or rS- configurations of QRS in V1 and V2, the presence of mid-QRS notching or slurring in at least two of leads V1, V2, V5, V6, I and aVL, and finally a QRS duration longer than 130 ms for women or longer than 140 ms for men. The detection algorithm is described in detail in [8].
- **NSIVCB** was detected when the QRS duration was longer than 120 ms and at the same time this record was not classified as LBBB or RBBB.
- **PR** was detected when the median of slope curtosis > 162 or sum of samples (where the sum of slope values > 140) > 2 or (number of possible spikes > 5 and median kurtosis of the signal with enhanced spikes > 250); exclusion of signals with artifact spikes; exclusion of signals where the sum of beats where the distance between spike and QRS is 0-300 ms is lower than half number of detected spikes.
- **LQRSV** was detected when maximum minus minimum of QRS amplitude (mean of this difference in all leads) is lower than 0.5 mV and at the same time maximum QRS amplitude in lead III is lower than 0.25 mV.

- **SNR** was determined when no other pathology was detected. Although the sinus rhythm can be present at the same time with other pathologies, in most databases this class is separated from the others.

Other pathologies are detected using more complex classification trees. Only CPSC, PTB-XL and Georgia database were used to train the classification tree. Five-fold cross-validation was used to avoid overfitting.

The most useful features for each pathology are described below. The features describing the QRS morphology were derived from the averaged QRS.

- **AF:** 1. ECG filtration with bandpass filter 20-40 Hz and then cross-correlation of sections 300 ms before the QRS onset; 2. Shannon entropy of RR intervals and NN intervals.
- **AFL:** 1. zero crossing of the 6th frequency band of stationary WT; 2. heart rate; 3. standard error of RR interval. AFL is detected only when AF is not detected in the same record.
- **SA** was detected using features obtained from the analysis of NN intervals. The NN interval indicates the distance between adjacent QRS complexes originating from the sinus node. The best features of our proposed features are defined by the following equations:
  max(NN)-min(NN);
  min(NN)/median(NN);
  max(NN) / median(NN).
  Another significant feature is heart rate. SA is detected only when AF, PR and AFL are not detected in the same record.
- **(C)RBBB:** 1. R duration in V1; 2. S duration in I, V1 and V6; 3. ID Time in II, III and V1; 4. PQ interval. ID Time is the time from QRS onset to the peak of the R.
- **IRBBB:** 1. T amplitude in V2-V6; 2. R duration in V1; 3. ID Time in V1.
- **LAnFB:** 1. maximum amplitude of QRS in II, III and aVF; 2. morphology of rS in II and aVF.
- **LAD:** 1. R duration in I and V6; 2. S duration in I; 3. maximum amplitude of QRS in I, II, aVF and V4.
- **RAD:** 1. maximum amplitude of QRS in I, III, V6; 2. S duration in V5.
- **PAC, SVPB:** The number of NN intervals shorter than 0.9 * median(NN) and at the same time the following NN interval is longer than 1.1 * median(NN)
- **PVC, VPB:** 1. minimum, maximum, variance, median and mean of correlation between aligned current QRS and pattern (median) QRS, 2. maximum number of leads in which the current QRS correlation coefficients are lower than 0.83, 3. minimum RR_{PVC} interval/median(RR).
- **TAb:** 1. T amplitude in aVR, V3-V6; 2. area under T in V3 and V6.

3. Results

Table 1. Classification success – F1-measure, Validation means challenge hidden validation dataset

	CPSC	CPSC 2	PTB-XL	Geor-gia	Validation
AF	0.9263	0.7394	0.9141	0.7528	0.83
AFL	-	0.0667	0.2626	0.0943	0.098
PR	-	0	0.9442	-	0
SA	-	0.0615	0.6057	0.5745	0.438
STach	-	0.8305	0.8114	0.8865	0.844
SB	-	0.1733	0.2754	0.9353	0.869
IAVB, LPR	0.7266	0.5571	0.5888 *	0.7212	0.696
RBBB, CRBBB	0.7262	0.7881	0.8385	0.7686	0.693
LBBB	0.7728	0.0506	0.8029	0.7432	0.697
IRBBB	-	0.2162	0.5994	0.4670	0.359
LAnFB	-	-	0.6674	0.3767	0.319
NSIVCB	-	0.0545	0.2486	0.1796	0.189
PAC, SVPB	0.6650	0.3651	0.4075	0.5473	0.557
PVC, VPB	-	0.4608	-	0.4531	0.28
LAD	-	-	0.6033	0.6920	0.63
RAD	-	0	0.6389	0.2471	0.246
Tab	-	0.0317	0.0235	0.2589	0.219
SNR	0.4092	0.0032	0.7172	0.5510	0.562
LQRSV	-	-	0	0.2679	0.261

* IAVB and LPR were merged into one group for PTB-XL

The success of the detection of individual pathologies from the four largest available databases can be seen in Table 1. These results were obtained on public parts of the databases (training score). Column validation means score for challenge hidden validation dataset, the

algorithm could be tested ten times on this data.

Some databases contain only a small number of some pathologies and therefore the success rates are unreliable. This is because the correct or incorrect classification of a few records will cause large changes in success. The values are highlighted in gray in Table 1 if the number of signals with the particular pathology in the database is less than 50. The number of individual pathologies in the validation set is unknown.

The final challenge validation score was 0.435, and full test score was 0.354, placing us 11 out of 41 in the official ranking.

4. Discussion

Detection of AF, STach, IAVB + LPR, RBBB, LBBB and LAD achieves consistently good success. PR pathology is represented mainly in the PTB-XL database, where the success rate is high. There are only three signals with PR in the CPSC 2 database, but none is detected correctly.

Classification success of SB is fluctuating. It is because some databases use a threshold of 60 bpm, while others use 50 bpm. The classification success of LAnFB and RAD is also fluctuating. This is probably due to the different definition of these pathologies in different databases.

Detection of AFL, NSIVCB, TAb and LQRSV has very little success. This is due to the vague definition of these pathologies and therefore poor-quality features. The vague definition of pathologies is difficult to solve with the procedure we use. The pathologies LQT, QAb, Brady, and TInv were excluded from the final version of the software due to the low success of the classification.

The advantage of our algorithm is that the algorithm is easy to interpret, it is obvious according to which features the signal is classified and what thresholds are used.

5. Conclusion

We introduced an algorithm for detection of 19 cardiac pathologies. The success of the proposed algorithm was determined on publicly available PhysionNet/CinC Challenge 2020 databases and on the hidden test set within the PhysioNet/CinC Challenge 2020 [1]. Our approach achieved a challenge validation score of 0.435, and full test score of 0.354, placing us 11 out of 41 in the official ranking, although 4 pathologies were not detected by our algorithm.

Acknowledgments

This work has been funded by the United States Office of Naval Research (ONR) Global, award number N62909-19-1-2006. The authors wish to thank LCDR Joshua Swift from ONR Code 342 and Dr. Stephen O'Regan from ONR Global Central and Eastern European Office for their support.

References

[1] Alday EAP, Gu, A., Shah A., et al., "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas.*

[2] Liu, F., Liu, C., Zhao, L., et al., "An open access database for evaluating the algorithms of ECG rhythm and morphology abnormal detection," *J. Med. Imaging Health Inform.*, vol. 8, no. 7, pp. 1368-1373, 2018.

[3] Goldberger, A., Amaral, L., Glass, L., et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circ.*, vol. 101, no. 23, pp. e215–e220, 2000.

[4] Bousseljot, R., Kreiseler, D., Schnabel, A., "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomed. Tech. (Berl)*, vol. 40, no. 1, 1995.

[5] Wagner, P., Strodthoff, N., Bousseljot, R., et al., "PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1)," *PhysioNet*, 2020.

[6] Wagner, P., Strodthoff, N., Bousseljot, R., et al., "PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1)," *Sci. Data*, vol. 7, Art. no. 154, 2020.

[7] Smital, L. Marsanova, L., Smisek, R., et. al., "Robust QRS detection using combination of three independent methods," *CinC*, vol. 47, Sep. 2020.

[8] Smisek, R., Viscor, I., Jurak, P., et al., "Fully automatic detection of strict left bundle branch block," *J. Electrocardiol.*, vol. 51, no. 6, pp. S31-S34, Nov.-Dec. 2018.

[9] Maršánová, L., Němcová, A., Smíšek, R., et al., "Advanced P wave detection in ECG signals during pathology: evaluation in different arrhythmia contexts," *Sci. Rep.*, vol. 9, Art. no. 19053, Dec. 2019.

Address for correspondence:

Radovan Smisek.
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology
Technická 3058/12, 61600 Brno, Czech Republic
E-mail address: smisek@feec.vutbr.cz