

Automatic Detection of Atrial Fibrillation and Other Arrhythmias in Holter ECG Recordings Using Rhythm Features and Neural Networks

Filip Plesinger, Petr Nejedly, Ivo Viscor, Josef Halamek, Pavel Jurak

Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

Abstract

Atrial fibrillation (AF) is a disease affecting 1-2 % of the population. Due to its episodic behavior, it is usually detected using Holter recordings. While various AF detection methods have been described in the past, it still remains problematic because holter recordings may contain other arrhythmias (OA) and, moreover, they may be influenced by patient movements. In accordance with the Physionet Challenge 2017, we propose an autonomous and robust method of distinguishing between pathological and normal recordings.

First, QRS complexes are identified using envelopograms (8-25 Hz and 70-90 Hz). Detected QRS complexes are clustered into morphology groups using a raw ECG signal. If too many morphology groups are produced or less than four QRS complexes are detected, the process is aborted and the recording is considered too noisy.

Next, a median shape for the first and the second morphological group of QRS complexes is built. Features are extracted from averaged shapes, from the rhythm of major morphology QRS complexes, from QRS correlation to template shapes and from the convolutional neural network. 277 features are fed into the neural network, resulting in three outputs. The 120 most important features, as well as outputs from the neural network, are fed into a bagged tree ensemble.

Machine-learning algorithms and logical rules were trained using 8,138 files from a reduced training set. The resultant F1 score measured using hidden test set (3,658 recordings) was 0.81 (normal 0.91, AF 0.80, OA 0.74).

1. Introduction

The task of the Physionet Challenge 2017[1] was to develop a method for automated classification of holter ECG recordings. Target classifications were normal (N), atrial fibrillation (AF), other arrhythmia (OA) and noisy (X).

A considerable number of methods for automated detection of atrial fibrillation have been presented in the past. The common weakness of these studies is, however, that they were not trained and tested on a public dataset or that they were focused on recognition between N and AF recordings only[1]. Moreover, these methods usually did not work with the noise that is usually present in ECG

holter recordings as a consequence of body movements or body presence in areas with a larger amount of electromagnetic noise.

Here, we present an automated method for ECG holter classification which is intended to work with regular ECG holter recordings (i.e. containing noise and movement artifacts).

2. Method

The presented method consists of the following blocks – signal transformation (mostly to envelopograms), QRS detection, signal averaging, feature extraction and their processing with machine learning (ML) and simple logical rules (Fig. 1). The processing is as follows: an ECG file (1 lead, 300 Hz sampling, AliveCor device) is loaded and transformed into envelopograms (Fig. 1B) intended for QRS detection (*LF*: 1-8 Hz, *MF*: 5-25 Hz, *HF*: 45-65 Hz) and for a convolution neural network (1-5 Hz, 5-10 Hz, etc. up to 35-40 Hz). Although the presented solution uses some features independent of QRS detection, most of them require detection of QRS complexes.

2.1 QRS detection and morphology groups

QRS detection (Fig. 1C) uses the idea of peak detection in subtracted $R = MF - HF$ envelopograms[2] which automatically removes part of the noise influence in ECG. Maxima in the R signal are detected and further tested in a 0.6 s. window (W) centered on the tested maxima. The test consists of three steps. First, peak amplitude must be at least 1.5x higher than any other value in W . Next, three consecutive segments (0.2 s.) of window W are defined as W_a , W_b and W_c . The standard deviation of R in segment W_b must be higher than the sum of standard deviations in segments W_a and W_c . Finally, the last test compares LF sums to the left and right of the tested maxima in 0.25 s. windows; the right LF sum must be higher than the left LF sum – this check reflects the expected presence of a T-wave. If the maximum passes these three checks it is considered a QRS complex. The following QRS complex cannot be closer than 175 milliseconds. If the total number of detected QRS complexes is lower than 4, the recording is considered noisy and the process is aborted (Fig. 1D,G).

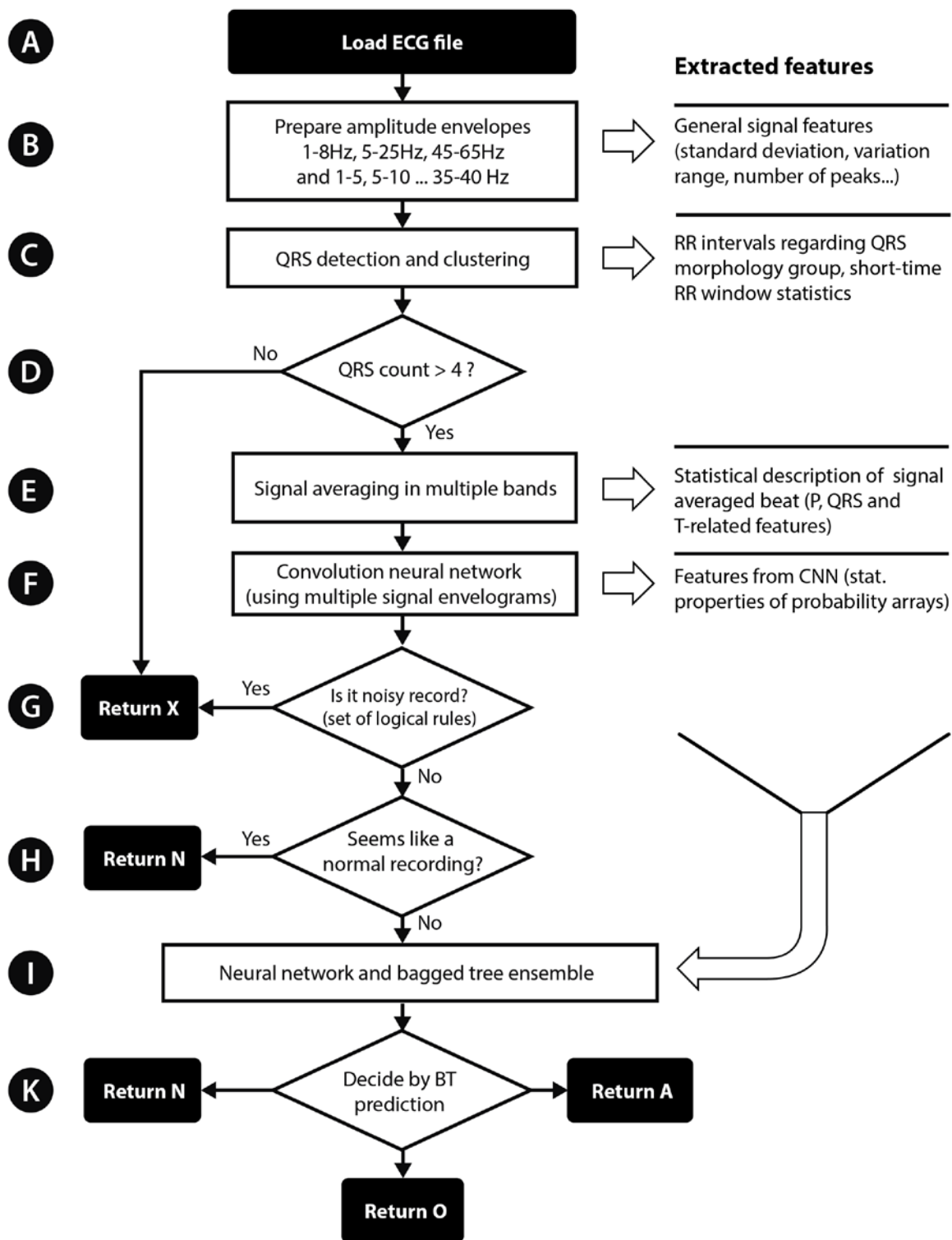


Figure 1: Flowchart of presented algorithm. The convolution neural network uses multiple envelopograms as input. Noisy recordings (X) are detected using features from QRS detection and convolution neural network outputs.

QRS complexes are clustered into groups by morphology. The first QRS complex defines the shape for the first group. If any other QRS correlates to this one (> 0.9), it is also assigned to the same group. If not, it creates a new morphological group (MG). If the number of created groups is higher than the length of the file in seconds, the file is considered noisy and the process is aborted. MGs are merged together if their averaged QRS correlates better than 0.95 when tested with time offset. Finally, MGs are ordered by the number of associated QRS complexes. Thus, typical healthy recordings contain only QRS associated to MG₁, while recordings with, for example, premature ventricular complexes (PVC) will contain a large number of sinus beats in MG₁ and PVCs in MG₂. QRS complexes associated with low-occurrence MGs usually belong to artifacts.

2.2 Feature extraction

Features were extracted from several sources. The first source was statistical description of RR intervals (prepared for MG₁ and MG₂ and also independent of MG) as RR standard deviation, often used for detection of atrial fibrillation (features extracted: $n=68$). Another source was statistical description of RR in a moving window (6 beats), also containing single RR crossings through average RR or description of RR distribution ($n=37$). Because detection of RR irregularities is essential for detection of other arrhythmias, irregularities in RR (Fig. 2-right) were tested as well as the ratio of *MF* and *LF* sums close to QRS.

Features extracted from RR intervals were used to identify noisy recordings – if RR intervals from MG₁₋₃ cover less than 40 % of the ECG signal, the signal was considered noisy.

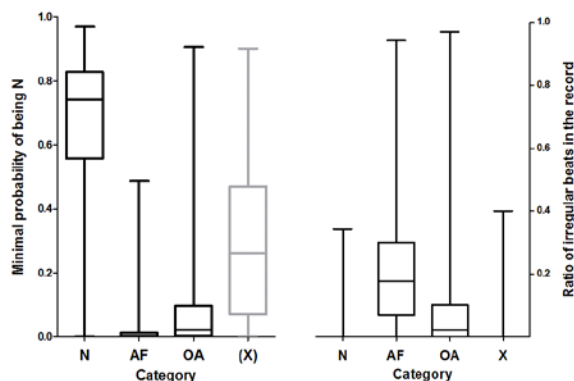


Figure 2: Examples of extracted features. Left – min. probability of being N from convolution neural network in 6-sec. moving window (noisy recordings were not predicted by CNN). Right – ratio of irreg. beats in RR intervals. N – normal, AF – atrial fib., OA – other arrhythmia, X – noisy. The Kruskal-Wallis test shows $p < 0.01$.

A significant source of features were averaged beats (Fig. 1E); averaging was performed using raw ECG and *LF* and *MF* signals. Only MG₁ beats (sinus beats) were used for signal averaging. Features describing amplitude and standard deviation in specific segments and their ratios were extracted. For example, standard deviation in a segment $<-200;-80>$ ms prior to QRS should be very low in recordings containing atrial fibrillation because the P-wave is not synchronized with QRS. Therefore, this segment in an averaged shape should be flat, while averaging of a healthy recording will reveal a P-wave even in a noisy signal. 60 features were extracted from averaged QRS, pointing to QRS, P-wave, T-wave and ST elevation.

Independent of QRS detection, a convolution neural network (CNN) was also used for feature extraction (Fig. 1F). The input for CNN was a floating 6-second window across band-passed (1-40 Hz) ECG and its envelopes (1-40 Hz by 5 Hz). The CNN contains 2 convolution layers, uses a ReLU non-linearity function [3] and contains 15 neurons in the first fully connected layer. The output layer contains 3 neurons and after normalization (softmaxLayer) the CNN produces the probability of states N, AF and OA (Fig. 2-left). Because we receive probabilities for each window, statistical descriptions of these probabilities (min, max, mean and standard deviation and their derivatives) were used as features ($n=17$).

Features were also extracted from correlation coefficients of average QRS (MG₁ and MG₂) to 30 shapes present in the training set ($n=62$), including other features related to correlation.

Existing features were used to isolate N recordings (Fig. 1H) using simple logical rules: BPM (derived from MG₁) must be between 50 and 99 BPM, max probability of OA derived from CNN must be lower than 0.35. This rule is capable of isolating 3,136 N recordings (63 % of all N) from the training set without following machine-learning techniques.

2.3 Processing with a neural network and bagged tree ensemble (BT)

All features ($n=277$) were fed into a neural network (Fig. 1I) with one input layer, one hidden layer (25 neurons) and one output layer (3 neurons). The output values of the NN (*i.e.* N, AF and OA probabilities) are added to the feature set and fed into a bagged tree (Fig. 1K). Finally, the output of the bagged tree is interpreted as an N, AF or OA state.

2.4. Training

The original training set contained 8,350 ECG recordings. We removed 167 of these recordings in cases

where we did not agree with the original classification. The modified training set (8,183 recordings) was then used to train the CNN and to extract features. Logical rules (Fig. 1G) identified 196 recordings as noisy (based on QRS and CNN features) and these records were not used for the training of the NN and BT. Finally, 277 features from 7,987 files were used to train the NN; their subset (120 features) was used to train the BT.

3. Results

The presented solution was tested on a hidden testing set containing 3,658 recordings. Training and testing scores are shown in Table 1.

Table 1. Results of the presented solution. BT – bagged tree.

	BT cross-validation (N=7,987)	Training set (N=8,183)	Testing set (N=3,658)
Normal	0.9529	0.96	0.91
Atrial fib.	0.8690	0.96	0.80
Other arr.	0.8575	0.92	0.74
Noisy	-	0.66	0.54
Total F1	89.32	0.95	0.81

4. Discussion

We experimented with different scenarios of feature collection and their processing with machine learning. The most straightforward variant of feature processing using simple logical rules was our primary approach, but it proved to be applicable only in specific cases (i.e. anything with $BPM \geq 100$ cannot be an N recording). No single extracted feature alone could be used for clustering into target groups with good accuracy. On the other hand, simple logical rules were used for detection of noisy recordings, because their count was too small for effective machine learning. Distinguishing the OA category – which may contain any other arrhythmia – from the AF category presented an extremely challenging task, especially because we were working with real holter ECG recordings. This led us to experiment with different kinds of machine-learning approaches (Table 2).

Solutions using an NN or BT or their combination as the last processing step were comparable (0.83 to 0.85) when they used the same feature set. The power of the solution using only CNN features (n=12) connected to an NN was, however, really surprising. This solution produced a test score of 0.80 which seems – considering the number of features used – to be an extremely powerful solution for recognizing N, OA and AF recordings. However, the best results were achieved with a combination of CNN features and all other features processed with an NN followed by a BT.

Table 2. Comparison of scores produced by different kinds of machine-learning methods and their combinations. ML – machine learning, NN – neural network, BT – bagged tree, CNN – convolutional neural network. Please note the score of 0.80 produced by CNN and NN alone without detection of QRS complexes. The training score with BT solutions is presented as cross-validation (after the slash). *Scores obtained after the end of the challenge with updated labels.

Used ML method	Feature based on		Result score for	
	QRS	CNN	training	testing
NN	-	Yes	0.89	0.80
NN	Yes	-	0.85	0.82
NN	Yes	Yes	0.89	0.84
BT	Yes	Yes	0.89	0.83
NN+BT	Yes	Yes	0.93*	0.81*

5. Conclusion

We presented a method for automated classification of holter ECG recordings into four groups: normal recordings, recordings with atrial fibrillation, recordings with any other arrhythmia, and noisy recordings. Our results demonstrated the necessity of using features based on QRS detection, but also showed the strength of features based on convolution neural networks.

Acknowledgements

The research was supported by Czech Academy of Sciences project MSM 100651602, by Czech Science Foundation grant GA17-13830S and by MEYS CR (LO1212).

References

- [1] G. Clifford, C. Liu, B. Moody, L. Lehman, I. Silva, Q. Li, A. Johnson, and R. Mark, “AF classification from a short single-lead ECG recording: The Physionet Computing in Cardiology Challenge 2017,” in *Computing in Cardiology*, 2017.
- [2] F. Plesinger, P. Klimes, J. Halamek, and P. Jurak, “Taming of the monitors: reducing false alarms in intensive care units,” *Physiol. Meas.*, vol. 37, no. 8, pp. 1313–1325, 2016.
- [3] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” *AISTATS '11 Proc. 14th Int. Conf. Artif. Intell. Stat.*, vol. 15, pp. 315–323, 2011.

Address for correspondence:

Filip Plesinger
 Institute of Scientific Instruments of the CAS
 Kralovopolska 147
 Brno 6012 64, Czech Republic
 E-mail address: fplesinger@isibrno.cz