

Classification of Atrial Fibrillation in Short-term ECG Recordings Using a Machine Learning Approach and Hybrid QRS Detection

Mateusz Soliński, Anna Perka, Jacek Rosiński, Michał Łepek, Joanna Rymko

Faculty of Physics, Warsaw University of Technology, Warsaw, Poland

Abstract

Atrial fibrillation (AF) is one of the most common sustained cardiac arrhythmia, occurring in 1-2% of the general population. Significant mortality and morbidity is related to occurrence of AF arrhythmia due to high risk of hospitalization, stroke, heart failure and coronary artery disease, etc. In many cases AF may not produce any symptoms and may go unnoticed by a patient, which is why there is a high importance to develop methods of detecting this heart disorder.

Creating an algorithm for AF and other arrhythmias classification of short-term single lead ECG signals was the aim of the PhysioNet Challenge 2017. The database was composed by over 8.5 thousand ECG recordings (between 10 sec and 60 sec length) measured by AliveCor device, provided by organizers.

We prepared an alternative hybrid approach for QRS detection in order to obtain RR time intervals. It consists of two complementary methods in hierarchical order: one based on nonlinear transformation and first-order Gaussian differentiator as superior and another one proposed in sample entry as inferior. We introduce the machine learning algorithm in order to classify whether it is normal sinus rhythm, AF or an alternative heart rhythm using features considered regularity of RR time intervals and morphology of the ECG signal. The separate part of the algorithm based on beat averaging method is dedicated for preceding extraction of too noisy recordings from the input to the classifier.

The best overall F1 score we achieved in the official phase of the PhysioNet Challenge 2017 was 0.77 (0.86 for normal, 0.78 for AF and 0.66 for other rhythms).

1. Introduction

Nowadays, digital technology allows to create various algorithms improving medical diagnostics. In cardiology, many of them are already implemented into devices both using in monitoring condition of a patient and in conducting therapy. Thus, the usage of such equipment become a significant enhancement in medicine.

We created method to differentiate AF rhythm from normal and other rhythm as it seems to be really common

arrhythmia in society [1]. Our approach is based on mostly intuitive features, that can be acquired even from dubious quality recordings of electrical activity of heart. Those features describe a rhythm, its irregularity, morphology of QRS complex and P-wave and diversity of those shapes.

The core of our algorithm is neural network with 78 input features. In order to improve proper classification of signals we decided to apply additional noise detection and preliminary other-signal detection which are used before the network. It allows us to exclude bad quality signals other signals with high specificity before further classification.

2. Data

This year's PhysioNet Challenge database contained 8528 ECG recordings collected using the AliveCor device. Time of signals was from 9 seconds to over 1 minute. The mean signal length was 30 seconds. Database contained 5154 signals classified as normal, 771 as AF, 2557 marked as other rhythm and 46 noisy (too noisy to be classified in one of the categories above). The test set to verify the algorithm effectiveness was unknown to competition participants and included 3658 ECG records of similar length as in the training base. The signals were bandpass filtered and sampled as 300 Hz by AliveCor device.

3. Classification algorithm

3.1. Preprocessing

The input of the algorithm are wavelet coefficients of ECG signal decomposition using Daubechies 6 wavelet. The preliminary analysis showed that the vectors of data based on d2 level wavelet coefficients of decomposition are more specific at classification process than usage exact ECG recordings.

3.2. Hybrid QRS detection

In order to create an algorithm for Atrial Fibrillation and other arrhythmias classification we extracted positions of R-peaks. It was the substantial part of the project because it allowed to obtain RR intervals. We prepared an alternative hybrid approach for QRS detection, which consists of two complementary methods in hierarchical order: One based on nonlinear transformation and first order Gaussian differentiator as superior and another one provided by organizers as inferior.

The given detector is taken from the Physionet Challenge in 2013 and it is based on Pan Tompkins algorithm with refractory period (between two R-peaks) of 250 ms [2]. There is a slight change in filtering process, because filtering process was done by using sombrero hat wavelet. The energy threshold was chosen at 98-99% of amplitude distribution. After the preprocessing operations, there is thresholding and search back stages which allow to find missed peaks by adaptivity of the detector.

The second detector is based on a nonlinear transformation and first order Gaussian differentiator [3]. The filtering here is done by usage of 15-th-order FIR filter using the least squares approach with two cut frequencies and they are respectively 6 and 20 Hz. An important phase of this algorithm is computing the Shannon energy. The outcome of this stage is filtered using the same filter as before. Then there is convolution with the first-order Gaussian differentiator, and the result is referred to as zero crossings function that has both positive and negative zero crossings because of the antisymmetric nature of the first-order Gaussian differentiator function. Negative zero crossings are detected and used as locations of R-peaks in the original ECG signal.

We have developed the hybrid detection in order to improve these methods and find as many peaks as possible.

3.3. Preliminary noise detection

A simple, preliminary noise detection algorithm was implemented before the exact signal classifier. The method was developed in aim to eliminate, from the main path of classification, the signals characterized by relatively high noise or lack of physiological data. It based on four features extracted either from ECG signal and RR interval time series and the threshold values determined for each of them:

A) similarity to averaged beat coefficient: each detected heart beat is segmented (90 samples from the left and 180 samples from the right) and normalized (by extracted the mean value and divided by standard deviation). Then the scalar dot products with averaged beat obtained from all detected beats in a signal are

calculated. The value of this feature is the number of beats which the dot product was under the value of 50 divided by the number of all detected beats. Threshold value: 0.5 (larger values classify the signal as 'too noisy').

B) beats density: the number of detected beats divided by the signal duration (in seconds). Threshold value: 0.68 (smaller values classify the signal as 'too noisy').

C) standard deviation of the QRS amplitudes (multiplied by 2). Threshold value: 3.2 (larger values classify the signal as 'too noisy').

D) number of zero-crossings of ECG signal differences divided by the number of all signal differences. Threshold value: 0.29 (larger values classify the signal as 'too noisy').

The signal is assigned as 'too noisy' and it is not evaluated in the next steps of algorithm if at least one of the feature indicates to high level of noise.

3.4. Basic other rhythm classification

The second preliminary classifier was developed to detect the other than sinus and AF rhythms in the most evident cases. It consists of two approaches: analyzing RR time intervals by using the heart rate and pNN50 (for bradycardia and tachycardia cases) and seeking for characteristic patterns indicated to ectopic, single atrial or ventricular preliminary beats or bigeminy and trigeminy. Each pattern consists of 4 successive RR intervals and it is characterized by specific relations between them. The signal is classified as 'other' rhythm when heart rate and pNN50 exceeds the threshold values or the appropriate number of ectopic patterns is detected.

3.5. Feature description

We used 78 features to classify ECG recordings in the neural network. We have few different families of features and they can be divided into 2 groups.

A) Based on RR intervals (63 features).

At the beginning, we re-engineered method provided by organizers [4], as it seems too sensitive to the length of recordings, achieving better results. In original method, there is parameter AFEvidence, which is calculated according to the equation:

$$\text{AFEvidence} = \text{IrregularityEvidence} - \text{OriginCount} - 2 * \text{PACEvidence},$$

where IrregularityEvidence is a sum of not empty bins from the Lorenz plot, OriginCount is a number of points near origin on Lorenz plot, and PACEvidence is a sum of non-empty bins in specific regions minus a number of points in this region.

As all of the parameters strongly depend on the number of points on the plot, we tried few different approaches. The best working case was when we took as features: numbers of points in different, specific regions of the chart and numbers of non-empty bins in those

regions. (We tried also a percentage of points in the specific region, but this approach worked little worse - we believe, that when you have longer recording, this approach could work better). We also extended the vector of RR intervals for recordings lasting 30 seconds or less. This led us to achieve alternative comparison of sums of points between recordings with extremely different length. By using this procedure, we obtained 25 features. 13 as Point count from 13 regions and 12 as not empty bins count in specific regions. Eight other features were extracted from the Poincare plot:

- percent of significant crossings through equal RR line;
- percent of crossings through equal RR line;
- one of the radiuses of an ellipse on the plot;
- second of the radiuses of an ellipse on the plot;
- pNNx, percent of decelerations between 2 consecutive RR intervals greater than x ms. We used values 50, 100, 75, 38 ms.

The term significant in that case means that the Euclidean distance was greater than 60 ms. Those features were similar to those extracted from Lorenz plot, but we think about them as more intuitive. Crossings through RR line correspond to acceleration of HR and next deceleration that often occur in AF rhythm.

Consequent five features were acquired through calculation of sample entropy from a vector of RR intervals. Template length was set to 5, the matching threshold as 0.2, we also used an option to standardize signal in order to compare results from different signals between them.

The next 25 features were generated as statistical characteristics of the RR intervals. These were simple features as: mean RR, standard deviation of RR, minimal RR, maximal RR, maximal gradient of normalized RR, minimal gradient of normalized RR and mean gradient of normalized RR, where normalization was performed by means of the mean value and standard deviation. We also take the mean of the three shortest intervals and the mean of the three longest intervals.

Further features were:

- 5-bin histogram of RR values normalized by means of the mean value;
- the number of zero crossings in the normalized RR;
- quantile of the normalized RR vector where three values were used alternately: 0.2, 0.5 and 0.7;
- the square root of the mean of the squares of the successive differences between intervals (RMSSD_x). If x stands for 1 then we take adjacent intervals. We also applied longer lags (x = 2, 3, 5);
- minimal autocorrelation in the RR intervals, where the correlation window length was the half of the RR vector length;
- maximal autocorrelation in the RR intervals;
- mean autocorrelation in the RR intervals.

B) Based on mean QRS morphology (15 features)

Other eight features were calculated from median QRS morphology. We aligned all QRS in the signal

considering 250 ms before R wave, and 350 ms after it occurs. Then we chose median of all amplitudes to achieve one QRS morphology, which won't be dependent on significant mistakes via wrong R peak detection (as a mean value would). Then we calculated following features:

- absolute value of field under plot (containing P wave);
- value of field under plot (containing P wave);
- sum of steepness of plot (containing P wave);
- a sum of a standard deviation of QRSs to Q wave (this parameter should answer if we have different morphologies);
- a sum of a standard deviation of QRSs to Q wave (this parameter should answer if we have different morphologies);
- steepness of R wave;
- zero crossings in median QRS;
- a value of field over an isoelectric line.

An isoelectric line is a line, which goes through 2 points. One at the beginning and this is a mean value of median QRS from beginning to point, which is 100 ms before R wave. The second point is point Q. To find this point we adjusted a line to the positive slope of R wave using least squares approach. The line was adjusted to points of mean QRS, beginning at 25 ms before R peak to one sample before R peak. There we also get steepness of R wave, as a parameter a of this line. If we know that, we can find point Q as a point on an adjusted line which value is closest to the isoelectric line. Connecting point one and two we get a line which is a close approximation of the isoelectric line and this allows us to calculate field over this line, as a parameter which can measure the presence of P wave.

We also did a little change in those features to see what kind of variability we can observe in long term mean QRS. To do so, we took median QRS similar parameters but considered as a long-term effect. We took 1 second before considered R wave and 1 second after it. Therefore, we could say something about variability in triplets of QRS complexes. We deleted only parameter R-steepness, and calculated other using the same method, and considered them as features, which could differentiate other rhythms and AF rhythm from Normal rhythm. That is how we get other seven features.

3.6. Algorithm structure and machine learning classification

The signal being processed is firstly analyzed with the hybrid detection algorithm and then the result is passed (without wavelet preprocessing) to the noise detection module. If the signal is not labeled as noise then the preliminary other-rhythm detection algorithm assesses the signal. If it is not assessed as other rhythm it is passed to the machine classification module where feature extraction and classification by the neural network occurs.

The decision algorithm structure is presented on the Fig. 1. The network classifier used in our algorithm is multilayer perceptron network with two hidden layers trained with the autoencoder technique in the MATLAB environment. The network architecture was optimized for the number of input features (78) and consists of the 75 units in the first hidden layer and 11 units in the second one. There are 3 output neurons corresponding to the output classes: normal, atrial fibrillation and other rhythm. The output with the highest value indicates the class of the sample.

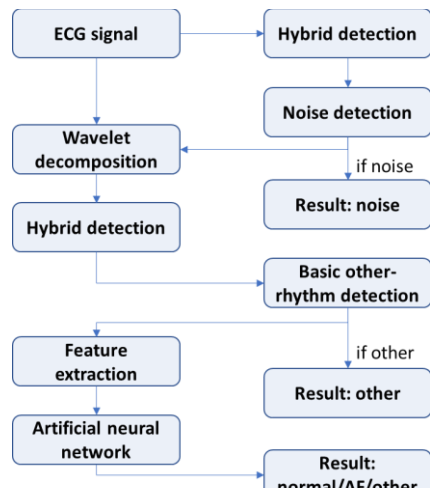


Figure 1. The decision algorithm structure.

Before the training of the network the dataset provided by the organizers is split into two subsets, the training set (75% of data) and the validation set (25% of data). As the numbers of recordings in our three classes are highly unequal, we duplicate the instances from the classes of lower numbers (other and AF) to have equal number of instances in each class and prevent biased learning. The instances in both training and validation sets are shuffled to obtain random order.

During the autoencoder training of the hidden layers we limit the training time to 1000 epochs and use scaled conjugate gradient algorithm and L2 weight regularization. Finally, both hidden layers and the output layer are stacked together to form one network and the fine tuning (final training) is performed. During the final training, we use Levenberg–Marquardt algorithm and limit training time to 12 epochs estimated basing on the mean squared error performance in time. The trained network is validated on the validation set. As the result of initialization, the network with random weights and the random shuffling the data in the training and validation sets, the value of recognition result of the single network can vary. To assess accurately the performance of the specific network architecture, we repeated the process of dividing the dataset, training autoencoders, final training and validation 10 times for each tested architecture. The network used in the final entry was chosen as the network of architecture of best performance with high scores for

both training and validation sets.

4. Results

Our final result in PhysioNet Challenge 2017 equaled 0.77 in overall F1 score. The F1 score of the signal classification as normal was 0.86, Atrial Fibrillation 0.78 and other rhythms 0.66.

5. Discussion & conclusions

The algorithm for classification of AF and other rhythms from short-term signal was developed. Apart from machine learning solutions the two preliminary simple methods for classification of evident cases of very noisy signals and easy to detect arrhythmias were used. They are characterized by relatively low sensitivity (70% for noise and 30% for other rhythms detection) but very high specificity (98% and noise and 99% for ‘other’ rhythms). Implementation of these methods increases F1 score for AF rhythm and other rhythms by 0.06 points (based on validation set of 300 signals).

One of the strengths of the algorithm was developed novel, hybrid method for QRS detection based on two popular algorithms and the hierarchical relationship between them. We noticed increases of F1 score for normal rhythm by 0.02 points, for AF rhythm by 0.07 point, for other rhythms by 0.05 points and for final F1 score by 0.04 points. Usage of neural network approach allowed to create the algorithm which is able to classify wide range of arrhythmias. We believe that it is a good solution for mobile diagnostic devices such as AliveCor.

References

- [1] Fuster V, Rydén LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heuzey JY, Kay GN, Lowe JE, Olsson SB. *ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation*, Europe, 2006, 8(9):651-745.
- [2] Behar J, Oster J, Clifford GD, *Non-invasive FECG extraction from a set of abdominal sensors*. In Computing in Cardiology Conference (CinC), 2013 Sep 22 (pp. 297-300). IEEE.
- [3] Kathirvel P, Prasanna SRM, Soman KP, An efficient R-peak detection based on new nonlinear transformation and first-order gaussian differentiator. *Cardiovasc Eng Technol* 2011;2(4):408-425.
- [4] Sarkar S, Ritscher D, Mehra R, A detector for a chronic implantable atrial tachyarrhythmia motor. *IEEE Trans Biomed Eng* 2008;55:1219-24.

Address for correspondence:

Mateusz Soliński
 Warsaw University of Technology, Faculty of Physics
 Koszykowa 75, 00-662, Warsaw, Poland,
 solinski@if.pw.edu.pl