Identification of Features for Machine Learning Analysis for Automatic Arrhythmogenic Event Classification

Vadim Gliner, Yael Yaniv

Biomedical Engineering Faculty, Technion-IIT, Haifa, Israel

Abstract

Cardiac arrhythmias are the leading cause of death in the western world, where atrial fibrillation (AF) is the most common arrhythmias. The PhysioNet/CinC 2017 Challenge aimed to trigger a design of an algorithm that accurately classifies short single ECG lead record to 4 categories: normal rhythm, atrial fibrillation, noisy segment or other arrhythmias. The algorithm was optimized on randomly selected records out of the challenge learning set (8528 records after reassuring it includes 60.43% of normal records, 0.54% of noisy records, 9.04% of AF records and 30% of other rhythm disturbance) and tested on hidden test database. A novel R peak detector was used to accurately detect the R peaks. Based on the R peak annotation, the P,Q,S and T peaks were detected and ECG beat morphology was extracted. Quadratic SVM classifier that include combination of 62 features was used to classify the short ECG record to one of the four categories mentioned above. For records which were classified as "normal" additional neural network classifier was applied.

Our algorithm reached results of total score (F_1) of 0.8 (ranked 24 out a total of 90 open-source software entries), whereas normal rhythm score (F_{1n}) was 0.9, AF rhythm score (F_{1a}) of 0.81, and other rhythm score (F_{1o}) of 0.69.

1. Introduction

Atrial fibrillation (AF) that is characterized by irregular and tachycardic heart rate, is the most common sustained cardiac rhythm disorder [1]. Although AF episodes per se are not dangerous, their side effects specifically increase in the risk for stroke are fatal [2]. Unfortunately, the available treatment today including invasive (catheter ablation) and non-invasive (drugs) are less than satisfactory [3].

Although, detection of AF is the first step to eliminate the side effect [4], AF detection and specifically automatic detection remains problematic, because it may be episodic. The classic methods for automatic detection of AF relate to the analysis of the absence of P waves. For that task neural network [5], statistical modelling [6] and wavelet analysis [7] were used. However, these methods perform poorly in the presence of noise, specifically from mobile device recordings. Later progress in the field relates to RR detection and the quantifying of its variability around the means (e.g., Poincaré plot analysis [8], entropy [9], etc.). Finally, recent approaches taking both approaches advantage by performing machine learning on the classification features [10] were reported.

More general approach was used recently to distinguish only between normal rhythm and other rhythm disturbance: arterial premature contraction, supraventricular tachycardia, premature ventricular contraction, ventricular tachycardia and ventricular fibrillation [11]. This approach used a support vector machine based classifier. Similar approach was used to distinguish arrhythmia in general from normal records [12]. However, in both cases no distinction was made between AF and other rhythms.

Despite the importance of automatic detection of AF three main factors eliminates the development of such algorithm: (i) the lack of gold standard databases from mobile ECG that include classification, (ii) robust algorithm that can accurately detect R peaks in the presence of AF episodes and noise and (iii) the ability to distinguish between AF and other source of arrhythmias.

The Physionet/Computing in Cardiology Challenge 2017 attempts to address the first limitation by providing classified data base of short single ECG lead record divided to 4 categories: normal rhythm, atrial fibrillation, noisy segment or other rhythm disturbance. Success of an algorithm to distinguish between the 4 categories will deal with the latter two challenges.

2. Methods

2.1. General

The PCinCC2017 consisted of 2 datasets; a training set of 8,528 single lead ECG recordings from 9s to just over 60s and test set contains 3,658 ECG recordings of similar length. All records consist of one bipolar channel recorded by AliveCor device. The data were sampled at 300 Hz and filtered by band pass filter in the device itself. The training set data was annotated to one of the four types: Normal (5154 out of 8528 records), AF (771 out of 8528 records), other rhythm (2557 out of 8528 records) and noisy (46 out of 8528 records). The scoring of the challenge uses a F_1 measure, which is an average of the three F_1 values from each classification type:

 ISSN: 2325-887X DOI:10.22489/CinC.2017.170-101

$$F_{1n} = \frac{2xN_n}{\sum N+\sum n} \quad (1)$$

where N is the total number of records the algorithm recognized as normal, N_n is the correct number of normal records classified by the algorithm as normal and n is real number of normal records.

$$F_{1a} = \frac{2xA_a}{\sum A+\sum a} \quad (2)$$

where A is the total number of records the algorithm recognized as AF, A_a is the correct number of AF records classified by the algorithm as AF and a is real number of AF records.

$$F_{1o} = \frac{2xO_o}{\sum O+\sum o} \quad (3)$$

where O is the total number of records the algorithm recognized as other rhythm, O_o is the correct number of other rhythm records classified by the algorithm as other rhythm and a is real number of other rhythm records.

2.2. R peak detection

We used our novel R peak detector that includes 7 mathematical manipulations (absolute value of Hilbert transform of the signal, first order derivative, polynomial fit of 34th order, Heuristic filter, second derivative of the signal, smoothing by convolution and a threshold). The algorithm deals with sudden movement of the patients, electrical drift, breathing noise, electrical spikes, environmental high frequency noise, reverse polarity, premature ventricular contraction, and enlarged P/T waves.

2.3. P, Q, S and T peaks detection

After R peaks were detected a segment of 0.3s before and after each R peak was defined as ECG beat. Averaging the defined ECG beats yielded the average ECG beat. After an average ECG beat was extracted, we found the first point after the R peak where the derivative changes sign and annotated it as S point. Similarly, the first point before the R peak where the derivative changes sign was annotated as Q point. T point was determined as the absolute maximal point between S point and S+100ms. P point was determined as absolute maximal between S-150ms to S point. Fig. 1 demonstrates the major ECG peak and beat morphology used here as features.

2.4. Learning strategy and algorithm

The training set (see above) was partially used for learning and part to test the algorithm. We divided the set to 7 parts, 6/7 was used for learning and 1/7 to test the algorithm performance. Each part includes 60.43% of normal records, 0.54% of noisy records, 9.04% of AF records and 30% of other rhythm disturbance. After learning on the first set, the data were mixed and again divided to 7.

A quadratic support vector machine (SVM) algorithm was used to make an initial classification into four categories mentioned above. The SVM was trained on 62 selected features, without performing principle component analysis first (PCA). Kernel function was quadratic and scaled. Multiclass method was "one-vs-one". After receiving the classifier results additional classification rules were applied to refine the decision algorithm. If the ECG record was classified by SVM as "Normal", additional neural network classifier was applied. This classifier was trained to distinguish only between "Normal" records and "Other" records. If the neural network classifier output was above 0.65 (probability of the record being "Other"), the record's classification was changed from "Normal" to "Other". This neural network classifier was a feed forward 2 layers network and consists of 20 neurons and was trained using the same 62 features. If less than 9 heart beats were detected, even if the algorithm classified the signal to either group, the algorithm classified it to noise. See Fig. 2 for the schema describing the algorithm steps.

2.5. Learning features

Features were calculated based on 6 categories: heart rate and heart rate variability [13], signal entropy [14], time-frequency domain features [15], intrabeat temporal interval variability [15], QRS morphology and average beat morphology [15]. We chose the features that provide the highest ability to distinguish between the groups, but, also ensure not to choose too many features which might lead to overfitting.

3. Results

In phase two, we obtained an overall F_1 score for the test set of 0.8. The F_1 scores for the different classes were 0.9, 0.81, 0.69 for normal, AF, and other respectively. See Table 1 for the average confusion matrix obtained during training validation.

We used 6 different categories of features for the learning algorithm. Time-frequency domain features were used to distinguish between other rhythm to other 3 categories. Heart rate and heart rate variability were used to distinguish mainly between AF and others. Signal entropy was used to distinguish between noise and other three categories. Intra-beat temporal interval variability was used to distinguish between other rhythm and the other 3 categories. QRS morphology was also used to distinguish mainly between other rhythm to the other 3 categories. Average beat morphology was used to extract features that are relevant mainly to other rhythm category, meaning arrhythmias that can be diagnosed by morphological abnormalities in either QRS segment or other beat parts.

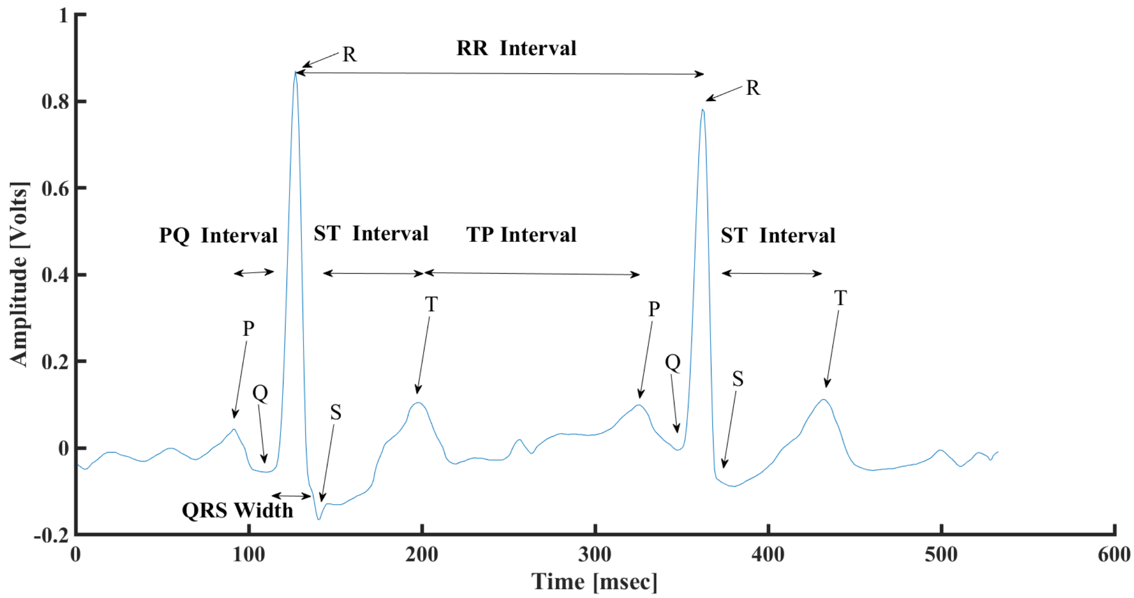

Figure 1: Major morphological features of ECG signal.

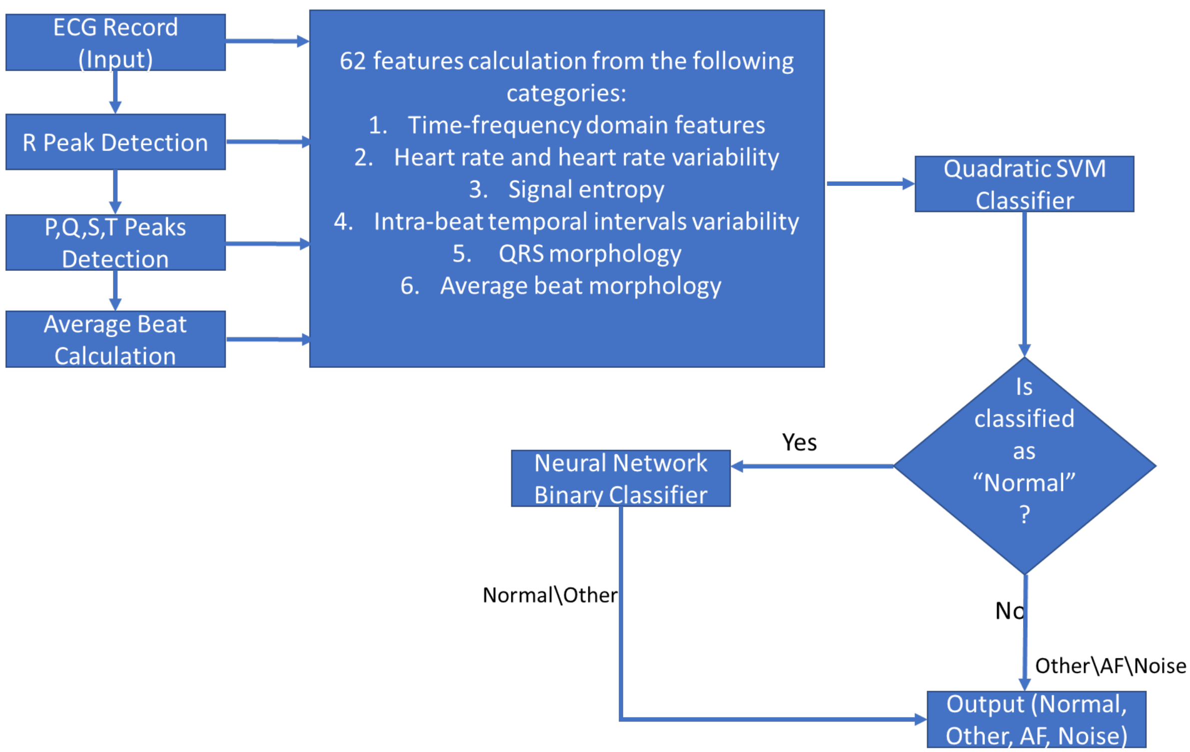

Figure 2: Block diagram of the algorithm to classify the ECG signal.

	Normal	AF	Other	Noise
Normal	4772	9	244	4
AF	17	650	70	0
Other	455	51	1960	10
Noisy	35	5	21	225

Table 1: Average confusion matrix obtained on the validation subset of the cross-fold for 7 cross-folds of the training set.

4. Discussion

We found here that 62 features were the optimal number of features for the learning strategy. Interestingly, less features would decrease F_1, but the same happen if more features were used. Thus, optimization of the number of features is not less important than choosing the features themselves.

The major element that allows us to calculate all the features used here is the precise R peak detector. Finding precisely the R peak allows to calculate the other waves and thus intra-beat temporal interval variability, QRS morphology and average beat morphology.

Acknowledgements

This work was supported by Max & Rachel Javit Fund in the Technion Autonomous Systems Program (TASP) (YY), Technion E.V.P.R. Star Fund (YY), Ilene and Steve Berger Fund (YY) and the Israel Science Foundation, No. 882/14 (Y.Y).

References

[1] G. Y. H. Lip *et al.*, “Atrial fibrillation,” *Nat. Rev. Dis. Prim.*, vol. 2, p. 16016, Mar. 2016.

[2] A. S. Go *et al.*, “Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study,” *Jama*, vol. 285, no. 18, pp. 2370–2375, 2001.

[3] A. Bajpai, I. Savelieva, and A. J. Camm, “Treatment of atrial fibrillation,” *Br. Med. Bull.*, vol. 88, no. 1, pp. 75–94, Dec. 2008.

[4] E. J. Benjamin *et al.*, “Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop,” *Circulation*, vol. 119, no. 4, pp. 606–618, 2009.

[5] A. Petrenas, V. Marozas, L. Sornmo, and A. Lukosevicius, “An Echo State Neural Network for QRST Cancellation During Atrial Fibrillation,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2950–2957, Oct. 2012.

[6] S. Ladavich and B. Ghoraani, “Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity,” *Biomed. Signal Process. Control*, vol. 18, no. 18, pp. 274–281, Apr. 2015.

[7] J. Ródenas, M. García, R. Alcaraz, and J. Rieta, “Wavelet Entropy Automatically Detects Episodes of Atrial Fibrillation from Single-Lead Electrocardiograms,” *Entropy*, vol. 17, no. 9, pp. 6179–6199, Sep. 2015.

[8] G. Tuboly and G. Kozmann, “Atrial Fibrillation Detection Based on Poincaré plot of RR Intervals,” *MEASUREMENT*, 2015.

[9] X. Zhou, H. Ding, W. Wu, Y. Zhang, N. Wessel, and P. Saparin, “A Real-Time Atrial Fibrillation Detection Algorithm Based on the Instantaneous State of Heart Rate,” *PLoS One*, vol. 10, no. 9, p. e0136544, Sep. 2015.

[10] “Machine Learning for Healthcare Technologies - The IET.” [Online]. Available: http://www.theiet.org/resources/books/healthcare/machlear.cfm. [Accessed: 12-Aug-2017].

[11] M. H. Song, J. Lee, S. P. Cho, K. J. Lee, and S. K. Yoo, “Support Vector Machine Based Arrhythmia Classification Using Reduced Features.”

[12] M. G. Tsipouras and D. I. Fotiadis, “Automatic arrhythmia detection based on time and time–frequency analysis of heart rate variability,” *Comput. Methods Programs Biomed.*, vol. 74, no. 2, pp. 95–108, May 2004.

[13] S. M. Pincus and A. L. Goldberger, “Physiological time-series analysis: what does regularity quantify?,” *Am J Physiol*, vol. 266, no. 4 Pt 2, pp. H1643-56, 1994.

[14] M. D. Costa, C. K. Peng, and A. L. Goldberger, “Multiscale analysis of heart rate dynamics: Entropy and time irreversibility measures,” *Cardiovasc. Eng.*, vol. 8, no. 2, pp. 88–93, 2008.

[15] V. Barbaro, P. Bartolini, G. Calcagnini, F. Martelli, and S. Morelli, “A comparison of methods for the classification of atrial fibrillation from intra-atrial electrograms,” in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, pp. 94–97.

Address for correspondence.

Yael Yaniv
Laboratory of Bioenergetic and Bioelectric Systems, Biomedical Engineering Faculty, Technion—IIT, Haifa, Israel
yaely@bm.technion.ac.il