Rhythm and Quality Classification from Short ECGs Recorded Using a Mobile Device

Joachim A. Behar[1], Aviv A. Rosenberg[1], Yael Yaniv[1], Julien Oster[2]

[1] Faculty of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa, Israel
[2] IADI, U947, INSERM, Université de Lorraine, Nancy, France

Abstract

Introduction: Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia. Its prevalence is 1-2% of the general population and it is associated with increased risk of mortality and morbidity.

Methods: The AliveCor mobile electrocardiogram (ECG) device was used to collect data. The Physionet Challenge aimed to create an intelligent algorithm for automated rhythm and quality classification. A database of 8528 single lead ECG was used for training and a closed database of 3658 ECG recordings was used for testing the participants algorithms on the Challenge server. The RR interval time-series was first estimated using a R-peak detector. Signal quality was estimated on a second-by-second basis and the continuous sub-segment with the highest quality was selected for further analysis. A number of features were estimated: heart rate variability (time domain based, fragmentation, coefficient of sample entropy etc.), ECG morphology (QRS length, QT interval etc.) and the presence of ectopic beats. The features were used to train support vector machine classifiers in a one-vs.-rest approach.

Results: For the final score of the challenge we obtained an overall F_1 measure on the test set of 0.80.

Conclusion: The feature based machine learning approach showed high performance in distinguishing between the different rhythms represented in the Challenge. This opens the horizon for computer automated interpretation of single lead mobile ECG.

1. Introduction

Atrial fibrillation (AF) is a cardiac pathology characterised by a chaotic contraction of the atrium, and is currently the most common cardiac rhythm disorder (1). The prevalence of AF roughly doubles with each advancing decade of age, from 0.5% at age 50-59 years to almost 9% at age 80-90 years (2). It is estimated that 2.3 millions adults in the US currently suffer from AF and that this figure will rise to 5.6 millions by 2050, reflecting the ageing of the population (3). AF is considered as a major cause for both mortality and morbidity as it increases the risk of heart failure, and stroke. AF is currently under-diagnosed, and is commonly detected only after a patient presents serious complications such as stroke or heart failure. Medication can ease symptoms and help preventing serious complications such as a stroke. Electrophysiological surgery and RF ablations have been shown to be quite effective treatments for restoring back a normal rhythm (4).

Recent progress in mobile technology (network, computational power, connectivity) makes it possible to develop low-cost, widely available and accurate medical devices. These devices can be used to address the shortage of healthcare resources in the developing world and lower the cost of healthcare in developed countries. AF is an excellent candidate for which the impact of such well-engineered mobile technology would be high. However, despite the availability of low-cost medical hardware, the ability to process data directly on the phone and the availability of large databases of biosignals there is yet very little that has been done in creating intelligent algorithms that could automatically interpret these medical data.

The subject of the Physionet/Computing in Cardiology challenge 2017 (5) addresses this topic and encourages researchers across the world to develop techniques for the classification of AF from a short single lead electrocardiogram (ECG) recording obtained using a mobile device.

2. Methods

2.1. Data

ECG recordings were collected using the AliveCor device and made available for the Challenge. An open database of 8528 single lead ECG and their annotations were used for training and a closed database of 3658 ECG recordings was used for testing on the Physionet Challenge server. Four categories of ECG recordings were present in the databases: atrial fibrillation (A), normal sinus rhythm (N), other rhythms (O) and noisy recordings ($\sim$).

2.2. Preprocessing

The RR interval time-series was first estimated using three different R-peak detectors, namely: $jqrs$ (6), $epltd$ (7) and $gqrs$ (8). The $jqrs$ detected R-peaks were used as the reference for RR interval estimation and the morphological analysis of the ECG. All three detectors were used to estimate the quality of a given record using $bsqi$ (9). It has been shown that AF detection accuracy decreases with a lowered ECG quality (10). In order to deal with transient noise, the segment of highest quality, defined as being the longest continuous segment with a minimal "second by second" $bsqi$ over a given threshold (taken to be 0.92).

2.3. Features

A set features were then extracted form the signal. These features can be divided in different categories : (i) signal quality, (ii) predictability of the RR intervals (11; 12), (iii) the ECG morphology (QRS duration, QT interval etc. (15)), (iv) heart rate variability (time based measures (13), fragmentation measures (13; 14). Figure 2 shows the distribution for a subset of the key features used on the different classes. A list of the subset of the features extracted is given in Table 1.

In particular, AF specific features can be divided in two categories: atrial activity based or ventricular based. The atrial activity based features include the analysis of the absence of P waves or the presence of f-waves ($ratio$, max_{freq} in Table 1). The ventricular based features are based on the predictability or the RR intervals ($CosEn$, AFE, OrC, IrE, $PACe$ in Table 1).

2.4. Machine learning

For classification we used a cascaded approach. This approach was chosen because of the large imbalance between the classes, the N class amounting for example to almost two thirds of all recordings in the training set. The features were used to train support vector machine (SVM) classifiers (16) with a Radial Basis Function (RBF) kernel in a one-vs.-rest approach. The first classifier was trained to distinguish between the N class against all other classes (O, A, $\sim$). The second SVM was trained to distinguish between A against O and the $\sim$ classes and finally the last SVM was trained to distinguish between $\sim$ and O. The classifiers were trained using repeated cross validation. Random search was performed to determine the SVM hyperparameters: soft margin constant C and RBF kernel hyperparameter γ. The features were normalised using a min/max approach and missing values were replaced by a default constant equal to -1.1, therefore emphasising the fact that missing values may carry information. The Feature selection was performed in stages, first the features were ranked by importance using the approach from (17). Secondly, a repeated cross-fold validation was performed with an increasing number of features in order to select the number of features to use in order to optimise the performances of the classifier while limiting the over-fitting.

3. Results

On the training set we obtained (see Table 2) an overall F_1 of 0.83. The F_1 measures for the different classes were 0.90, 0.83, 0.75 for normal, AF and other respectively. In phase two, we obtained an overall F_1 measure on the test set of 0.80. The F_1 measures for the different classes were 0.89, 0.82, 0.70 for normal, AF and other respectively. Our final score evaluated on the whole corrected test set provided at the end of the challenge was 0.80 which is consistent with our phase 2 results.

4. Discussion

Table 3 contains the average confusion matrix obtained during repeated cross-fold validation for the entry 3 of phase 2. In particular, there is a relatively high number of On (533) and No (376), which explains the relative low score for $F_{1,O}$ compared to the scores obtained for the classification of N and A recordings. A better understanding of the source for these misclassifations should improve the overall performance.

Figure 1 shows the influence of adding features to the SVM and the phenomenon of over-fitting (i.e. when training results are improving whereas validation results are decreasing).

An analysis of the ranking indicates that different features were selected for each SVM. The top five features for distinguishing the N class against all other classes were: min_{rr}, AVNN, PAS, medR and medST. The top five features for distinguishing the A class against O and $\sim$ were: IrE, CosEn, AFE, F_1, $median_{rr}$. The top five features for distinguishing the $\sim$ class against the O class were: medTamp, PACe, stdP, F_1 and stdR. In particular, for the A class, the ventricular features performed better than the atrial activity based ones. This is because the ventricular features are more robust to the presence of noise, which is particularly relevant for the Challenge databases (non-contact portable ECGs). Figure 2 illustrates the per class distributions for two of the most relevant features.

5. Conclusion

Overall, the feature-based machine learning approach showed good performance in distinguishing between the different rhythms available, with particularly high statistics for the detection of AF recordings.

Features	Definition
bSQI	Signal quality of the overall recording (9)
CosEn	Coefficient of sample entropy (11)
AFE	AFEvidence (12)
OrC	Number of points in the bin containing the Origin (12)
IrE	Irregularity Evidence (12)
PACe	PAC Evidence (12)
min_{rr}	Minimum RR interval
max_{rr}	Maximal RR interval
$median_{rr}$	Median RR interval
$nb_{outliers}$	RR-interval outliers. An outlier was defined as a sample exceeding 20% of a window average of size 12 beats.
medR	Median R-peak amplitude (mV)
stdR	Standard deviation of the R-peak amplitude (mV)
medQT	Median distance from Q_{on} to T_{off}.
$medQT_b$	Median QT interval corrected using the Bazett's formula
$medQT_{fre}$	Median QT interval corrected using the Frederica's formula
$medQT_{fra}$	Median QT interval corrected using the Framingham formula
$medQT_{hod}$	Median QT interval corrected using the Hodge formula
medQS	Median QRS interval length
stdQS	Standard deviation of the QRS intervals
medP	Median P-wave length defined as the distance from P_{on} to P_{off}
stdP	Standard deviation of the P-wave length
medPR	Median PR interval defined as the distance from P_{on} to Q_{on}
stdPR	Standard deviation of the PR interval
medPamp	Median P-wave amplitude defined as the amplitude of the P-wave computed from P_{off} to the peak of the P-wave.
medPRseg	Median PR segment defined as the distance from P_{off} to Q_{on}
medT	Median T-wave length defined as the distance from T_{on} to T_{off}
stdT	Standard deviation of the T-wave length
medTamp	Median T amplitude computed as the amplitude in mV between the T_{off} to the peak of the T-wave.
stdTamp	Standard deviation of the T-wave amplitude
medST	Median segment defined as the distance between QRS_{off} and T_{on}
medSTvar1/2	Amplitude of the ST segment defined as the median amplitude between the PR segment and the ST segment. The amplitude of the PR segment is computed as the median amplitude between P_{off} to QRS_{on}. The amplitude of the ST segment is computed as the median amplitude between the QRS_{off} and T_{on} or alternatively QRS_{off}+60 ms to the T_{on}.
AVNN	Average NN interval duration (ms) (13)
SDNN	Standard deviation of NN interval duration (ms) (13)
RMSSD	Root-mean-squared difference between adjacent NN intervals (ms) (13)
pNN50	Percent of NN interval differences greater than 50 milliseconds (%) (13)
SEM	Standard error of the mean NN interval (ms) (13; 14)
PIP	Percentage of inflection points (%) (13; 14)
IALS	Inverse average length of segments (13; 14)
PSS	Percentage of NN intervals that are in short segments (13; 14)
PAS	Percentage of NN intervals that are in alternation segments of at least 4 intervals (%) (13; 14)
ratio	Ratio of the power spectral frequency in the band 5-9 Hz normalised by the total power frequency computed on the PQRST cancelled signal
max_{freq}	Peak frequency in the band 4-45 Hz from the power spectrum computed on the PQRST cancelled signal

Table 1: Subset of the features extracted for each recording and used within the SVM classifier to distinguish between the different classes (N/A/O/∼). All the morphological fiducials were computed using the wave detector from (15).
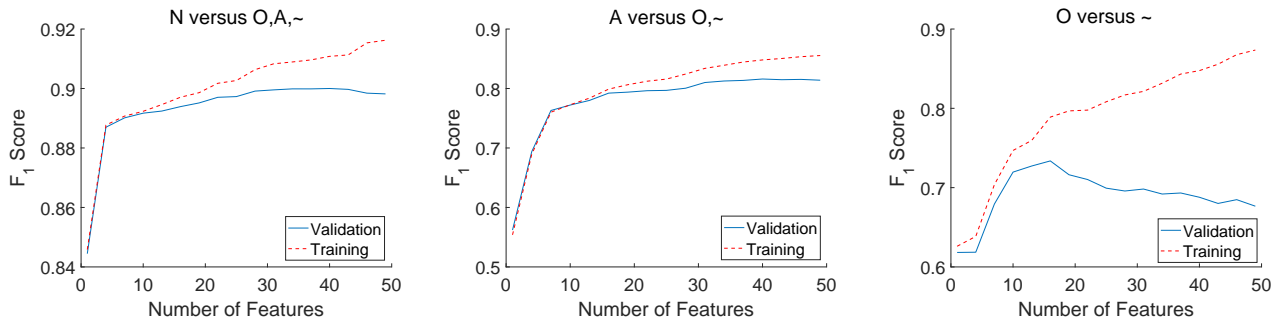
Figure 1: F_1 score obtained on the averaged training and validation sets during cross-validation performed by the SVMs trained for N versus O,A,∼ and A versus O, ∼ and O versus ∼. The figure illustrates the impact of adding features to the classifier in getting better average performance on both training and validation sets. For producing the plot the feature importance are first ranked and then added one by one for training and validating the SVM.
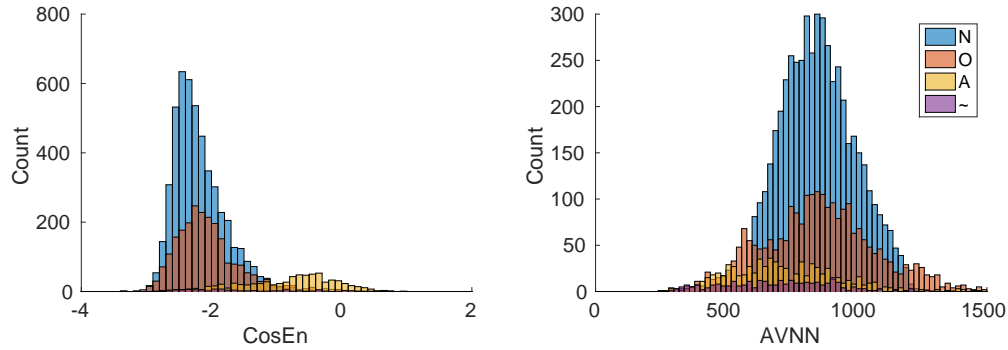
Figure 2: Per class distributions for two important features (AVNN and CosEn).

	Results	
	RCF	TS
$F_{1,N}$	0.90	0.89
$F_{1,A}$	0.83	0.82
$F_{1,O}$	0.75	0.70
$F_{1,\sim}$	0.65	NA
F_1	0.83	**0.80**

Table 2: Results obtained on the repeated cross-fold (RCF) validation sets and on the test set (TS, phase-2). NA: not available.

	n	a	o	$\sim$
N	4618	23	376	33
A	20	605	108	5
O	533	93	1792	39
$\sim$	50	7	52	175

Table 3: Average confusion matrix obtained on the validation subset of the cross-fold for the entry 3 of phase 2.

Acknowledgements

References

[1] G.Y.H. L, Fauchier L, Freedman S, Van Gelder I, Natale A, Gianni C, Nattel S, Potpara T, Rienstra M, Tse H, Lane D. Atrial fibrillation. Nature Reviews Disease Primers 2 2016;(3):16016.

[2] National Collaborating Centre for Chronic Conditions (Great Britain). Atrial fibrillation: national clinical guideline for management in primary and secondary care. Royal College of Physicians, 2006; .

[3] Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE. 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 2001;285(18):2370–2375.

[4] Calkins H, Kuck KH, Cappato R, Brugada J, Camm AJ, Chen SA, Crijns HJ, Damiano Jr RJ, Davies DW, DiMarco J, et al. 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. Europace 2012; 14(4):528–606.

[5] Clifford G, Liu C, Moody B, Silva I, Li Q, Johnson A, Mark R. AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017. Computing in Cardiology Rennes 2017;44.

[6] Behar J, Johnson A, Clifford GD, Oster J. A comparison of single channel fetal ecg extraction methods. Annals of biomedical engineering 2014;42(6):1340–1353.

[7] Pan J, Tompkins WJ. A real-time qrs detection algorithm. IEEE transactions on biomedical engineering 1985;BME-32(3):230–236.

[8] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit, and physionet. Circulation 2000; 101(23):e215–e220.

[9] Behar J, Oster J, Li Q, Clifford GD. Ecg signal quality during arrhythmia and its application to false alarm reduction. IEEE transactions on biomedical engineering 2013;60(6):1660–1666.

[10] Oster J, Clifford GD. Impact of the presence of noise on rr interval-based atrial fibrillation detection. Journal of electrocardiology 2015;48(6):947–951.

[11] Lake DE, Moorman JR. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. American Journal of Physiology Heart and Circulatory Physiology 2011;300(1):H319–H325.

[12] Sarkar S, Ritscher D, Mehra R. A detector for a chronic implantable atrial tachyarrhythmia monitor. IEEE Transactions on Biomedical Engineering 2008;55(3):1219–1224.

[13] Rosenberg AA. Non-invasive analysis of pacemaker function: Using heart rate variability to reveal the clock-like pacemaker mechanisms of the sino-atrial node. MSc Thesis Technion IIT 2017;.

[14] Costa MD, Davis RB, Goldberger AL. Heart rate fragmentation: A new approach to the analysis of cardiac interbeat interval dynamics. Frontiers in physiology 2017;8.

[15] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ecg delineator: evaluation on standard databases. IEEE Transactions on biomedical engineering 2004;51(4):570–581.

[16] Chang CC, Lin CJ. Libsvm: a library for support vector machines. ACM transactions on intelligent systems and technology TIST 2011;2(3):27.

[17] Yan K, Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators B Chemical 2015;212:353–363.

Address for correspondence:

Joachim A. Behar, PhD
Technion-IIT, Haifa, Israel
jbehar@technion.ac.il