

Multi-source Features and Support Vector Machine for Heart Rhythm Classification

Chengyu Liu^{1,2}, Qiao Li², Pradyumna B. Suresh³, Adriana Vest², Gari D. Clifford^{2,4}

¹School of Instrument Science and Engineering, Southeast University, Nanjing, China

²Department of Biomedical Informatics, Emory University, Atlanta, USA

³School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA

⁴Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, USA

Abstract

The PhysioNet/CinC Challenge 2017 provides over 8,528 short single channel ECG recordings for the classification of rhythms as normal sinus rhythm, AF, other rhythm, or too noisy. We present a support vector machine (SVM)-based heart rhythm classifier that leverages features based on rhythm, morphology and arrhythmia characteristics of the ECG. Four benchmark signal quality features, eleven rhythm features, fifteen morphology features, six arrhythmia features and four novel abnormality indices, were presented to the libSVM software package with 10 fold cross-validation. The SVM parameters were tuned using the modified cuckoo search (MCS) algorithm during the training.

Since two of the authors were involved in the challenge, this is an unofficial entry. Without the MCS parameter optimization, the results for the mean F_1 measures from the 10 fold cross validation on the training set were 0.911, 0.826, 0.728 and 0.852 for normal sinus rhythm, AF rhythm, other rhythm and noise respectively, resulting in a F_1 score of 0.822. With MCS parameter optimization, the mean F_1 measures on the training set increased to 0.921, 0.835, 0.739 and 0.870 respectively, resulting in a F_1 score of 0.832. The final F_1 measures on the test set were 0.914, 0.805, 0.691 and 0.737 for normal sinus rhythm, AF rhythm, other rhythm and noise respectively, resulting in a final F_1 score of 0.803.

1. Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, occurring in 1-2% of the general population and is associated with significant mortality and morbidity [1; 2]. Accurate detection for AF, especially from a variety of different arrhythmia types, is still challenging since some other arrhythmia types have the similar rhythm with AF. The PhysioNet/CinC Challenge

2017 provides a chance for the classification of rhythms as normal sinus rhythm, AF, other rhythm, or too noisy, from a short single channel ECG recordings. The details about the challenge can refer to [3].

In this study, we proposed a support vector machine (SVM)-based heart rhythm classifier that leverages features based on timing, morphology and spectral characteristics of the ECG, aiming to see how well when performing the common multi-source features and SVM for AF classification. We tried to use as many types of different AF features as possible since the previous studies confirmed that the multi-source features based machine learning approaches can provide an enhanced performance on AF detection [4-6].

2. Methods

2.1. Data

Data to support development and evaluation of challenge entries were collected by AliveCor device and provided freely via PhysioNet/CinC Challenge 2017 [3]. The data set includes 12,186 single lead ECG recordings lasting from 9 s to just over 60 s, with a sample rate of 300 Hz. The opened training set contains 8,528 recordings and the hidden test set (3,658 recordings) is unavailable to the public. Each recording was labeled as one of four types: normal sinus rhythm, AF, an alternative rhythm, or is too noisy to be classified. The labels were initially annotated by AliveCor device and then a few of them were manually corrected by the challenge organizers and experts. Detailed data profile can refer to [3].

2.2. Method description

Figure 1 shows the block diagram of the proposed SVM and multiple feature-based method, which consisted of four steps:

- Step 1: signal quality assessment for each 10 s

ECG episode.

- Step 2: feature calculation for the ECG episodes with good signal quality. Detailed feature definitions are given in section 2.3.
- Step 3: SVM (libsvm) classification using default parameter setting.
- Step 4: SVM (libsvm) classification using optimized parameter setting by using the modified cuckoo search algorithm [7; 8].

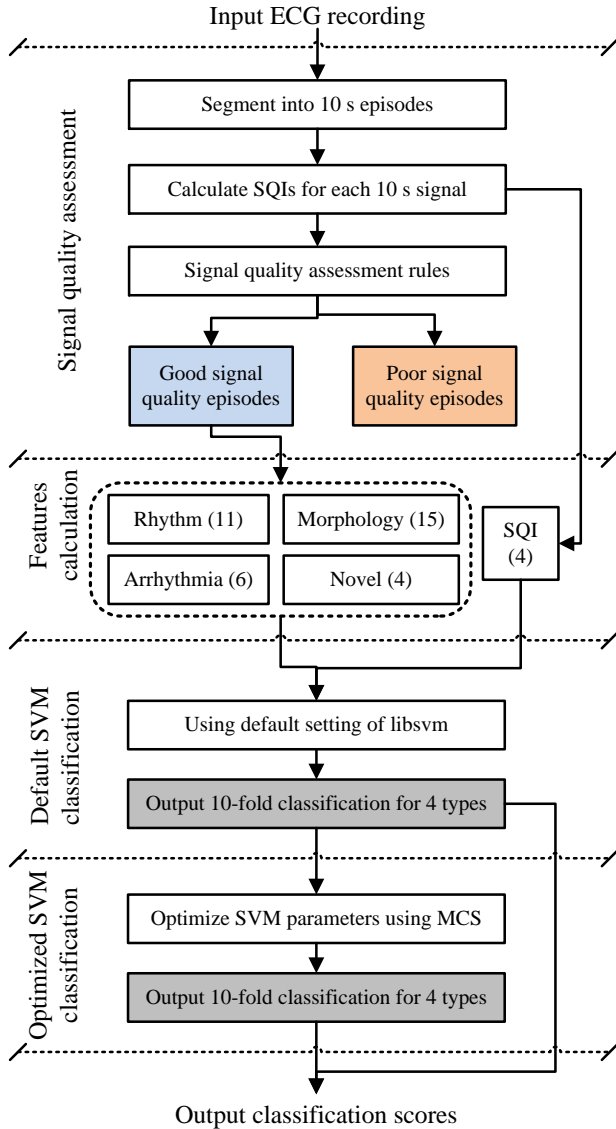


Figure 1. Block diagram of the proposed SVM and multiple feature-based method.

2.3. Features

We considered the features from five information sources. Signal quality index (SQI) was used as the first information source. The information about the heart rhythm, waveform morphology for QRS complex, P wave

and T wave, if arrhythmia is presented, and the new developed novel features were used as the other four information sources.

A. Signal quality features (4) [9; 10]

- bsQI: percentage of beats detected by ‘jqrs’ that were also detected by Pan&Tompkins method.
- sSQI: third moment (skewness) of the waveform distribution.
- ksQI: fourth moment (kurtosis) of the waveform distribution.
- fsQI: ratio of power P(5-14 Hz)/P(5-40 Hz) from spectrum analysis of burg method.

B. Rhythm features (11)

- MAD: thresholding on the median absolute deviation of RR intervals [11].
- AFEv: AF feature based on the analysis of Lorenz plot analysis [12].
- NFEn: normalized fuzzy entropy of the RR interval time series [13].
- noRR: number of RR intervals.
- mRR: mean value of RR intervals.
- minRR: minimum value of RR intervals.
- maxRR: maximum value of RR intervals.
- medHR: median value of heart rate calculate from each RR interval.
- SDNN: standard deviation of RR intervals.
- PNN50: mean value of RR intervals.
- RMSSD: root-mean square differences of RR intervals.

C. Morphology features for QRS complexes (5)

- mAmp: mean value of amplitudes of QRS complexes.
- sdAmp: standard deviation of amplitudes of QRS complexes.
- rAmp: ratio of 85% and 15% fractiles of the sorted amplitudes of QRS complexes.
- scovMor: correlation coefficient of QRS complexes in a small window (200 ms).
- lcovMor: correlation coefficient of QRS complexes in a large window (400 ms).

D. Morphology features for P and T waves (10)

- mAmpP & sdAmpP: mean and SD of P wave amplitudes
- mP2QRS: mean of P amplitude/QRS amplitude
- sdP2QRS: SD of P amplitude/QRS amplitude
- mAmpT & sdAmpT: mean and SD of T wave amplitudes
- mT2QRS: mean of T amplitude/QRS amplitude
- sdT2QRS: SD of T amplitude/QRS amplitude
- QT & QTc: QT and calibrated QT interval

E. Arrhythmia features (6)

- PAC1-3: three PAC prediction indices based on impulse rejection filters
- PVC1-3: three PVC prediction indices based on frequency domain filters

F. Novel features (4)

- MAI: morphology abnormal index for single beat ECG waveforms.
- TAI: template abnormal index for QRS complexes.
- AAI: amplitude abnormal index for single beat ECG waveforms.
- IAI: interval abnormal index for RR interval time series.

2.4. SVM classifier

The matrix of the forty features for the 8,528 training set recordings was calculated. We used the open-source libsvm software package to train the feature matrix and learn the SVM models [14]. Ten-fold cross-validation method was used to enhance the generalization ability of the trained SVM models. The default parameter setting for training SVM models were used as: kernel function: radial basis function; gamma parameter γ in kernel function: $1/N_{\text{feature}}$ ($N_{\text{feature}}=40$ here); cost parameter C : 1; weights for parameters C in each of the four classification type: 1.

Then, we kept using the 10-fold cross validation and using the kernel function of SVM as radial basis function, and tuned other SVM parameters using the modified cuckoo search (MCS) algorithm [7; 8]. The objective function was set to maximize the mean F_1 measure for the first three rhythm types (normal, AF and other rhythm).

2.5. Scoring mechanism

The scoring for this challenge uses a F_1 measure. For each of the four classification type, F_1 is defined as:

- Normal rhythm: $F_{1n} = \frac{2 \times Nn}{\Sigma N + \Sigma n}$
- AF rhythm: $F_{1a} = \frac{2 \times Aa}{\Sigma A + \Sigma a}$
- Other rhythm: $F_{1o} = \frac{2 \times Oo}{\Sigma O + \Sigma o}$

- Noisy: $F_{1p} = \frac{2 \times Pp}{\Sigma P + \Sigma p}$

The final challenge score was generated as [3]:

$$F_1 = \frac{F_{1n} + F_{1a} + F_{1o} + F_{1p}}{4} \quad (1)$$

3. Results

3.1. SVM classification using default parameter setting

The left half of the Table 1 shows the results from the 10-fold cross validation for the training set when using the default parameter setting for training SVM model. Mean F_1 results are 0.911, 0.826, 0.728 and 0.852 for normal sinus rhythm, AF rhythm, other rhythm and noisy types respectively, resulting in a mean F_1 of 0.822. Normal type hold the largest F_1 measure of 0.911, followed by the noisy class of 0.852, and the AF type of 0.826. Other type had the lowest largest F_1 value of 0.728. The standard deviation (SD) values of the F_1 results for the four types are 0.019, 0.045, 0.026 and 0.027 respectively. The total SD for the mean F_1 results from the 10-fold cross validation is 0.020.

3.2. SVM classification using MCS optimized parameter setting

Results for the training set using the optimized SVM parameters were shown in the right half of the Table 1. Mean F_1 results from the 10-fold cross validation are 0.921, 0.835, 0.739 and 0.870 for normal sinus rhythm, AF rhythm, other rhythm and noisy types respectively, resulting in a mean F_1 of 0.832. Compared with using the default SVM parameter setting, using the optimized SVM parameter enhanced the mean F_1 results for each of the four classification types, by 0.010, 0.009, 0.011 and 0.018 respectively. The SD values of the F_1 results are 0.012,

Table 1. Results of the multi-source features and SVM-based algorithm on training (10-fold cross validation) and test sets.

Set	Fold	SVM with default parameters					SVM with MCS optimized parameters				
		Normal	AF	Other	Noisy	Total	Normal	AF	Other	Noisy	Total
Training	1	0.878	0.746	0.690	0.845	0.771	0.905	0.746	0.705	0.879	0.785
	2	0.910	0.830	0.723	0.808	0.820	0.914	0.836	0.726	0.835	0.825
	3	0.933	0.783	0.738	0.843	0.818	0.933	0.812	0.763	0.855	0.836
	4	0.899	0.871	0.736	0.854	0.835	0.907	0.860	0.733	0.864	0.833
	5	0.929	0.844	0.708	0.861	0.827	0.922	0.834	0.730	0.871	0.829
	6	0.925	0.777	0.753	0.856	0.818	0.920	0.773	0.750	0.859	0.814
	7	0.905	0.875	0.762	0.864	0.847	0.940	0.895	0.768	0.877	0.867
	8	0.886	0.841	0.748	0.908	0.825	0.922	0.866	0.767	0.922	0.852
	9	0.923	0.878	0.687	0.861	0.829	0.940	0.898	0.714	0.890	0.851
	10	0.924	0.818	0.741	0.817	0.828	0.915	0.829	0.733	0.845	0.826
		Mean	0.911	0.826	0.728	0.852	0.822	0.921	0.835	0.739	0.870
	SD	0.019	0.045	0.026	0.027	0.020	0.012	0.045	0.022	0.025	0.023
Test							0.914	0.805	0.691	0.737	0.803

0.045, 0.022 and 0.025 respectively for the four types. The total SD for the mean F_1 results from the 10-fold cross validation is 0.023.

As shown in Table 1, the final F_1 measures on the test set were 0.914, 0.805, 0.691 and 0.737 for normal sinus rhythm, AF rhythm, other rhythm and noise respectively, resulting in a final F_1 score of 0.803.

4. Discussion and conclusion

In this study, we present a support vector machine (SVM)-based heart rhythm classifier that leverages features based on rhythm, morphology and arrhythmia characteristics of the ECG. Features from benchmark signal quality, RR rhythm, QRS/P wave/T wave morphology, arrhythmia and novel abnormality indices were employed. The open-source libSVM software package with 10 fold cross-validation was used to test and validate the usefulness of the multiple sourced features. In addition, we tested the effect of tuning the SVM parameters by the modified cuckoo search (MCS) algorithm on the classification accuracy. Using the MCS parameter optimization, improved mean F_1 measure of 0.832 on the training set from the 10 fold cross validation was achieved. As comparison, without MCS parameter optimization, the mean F_1 measure on the training set from the 10 fold cross validation was 0.822. When tested on the test set, a final F_1 score of 0.803 was achieved. Please note that we are unofficial entry since two of the authors were involved in the challenge.

The current study used multiple sourced features for the four classification tasks. We expect that, more features, including specific design for other rhythm type, VT/VF detection, spectrum/cepstral analysis, *etc.*, could further enhance the classification performance. In addition, no feature selection step was used in the current study. Thus, in the next phase, selecting more efficient features and excluding the redundancy and relevance between features, could also be helpful to enhance the classification performance.

Acknowledgements

Chengyu Liu would like to thank the funding from the International Postdoctoral Exchange Programme of the National Postdoctoral Management Committee of China.

References

- [1] Lip G Y H, Fauchier L, Freedman S B, Van Gelder I, Natale A, Gianni C, Nattel S, Potpara T, Rienstra M, Tse H, Lane D A. Atrial fibrillation. *Nature Reviews Disease Primers*. 2016; 2: 16016.
- [2] Camm A J, Kirchhof P, Lip G Y, Schotten U, Savelieva I, Ernst S, Van Gelder I C, Al-Attar N, Hindricks G, Prendergast B, Heidbuchel H, Alfieri O, Angelini A, Atar D, Colonna P, De Caterina R, De Sutter J, Goette A, Gorenk B, Heldal M, Hohloser S H, Kolh P, Le Heuzey J Y, Ponikowski P, Rutten F H. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J*. 2010; 31: 2369-2429.
- [3] Clifford G D, Liu C Y, Moody B, Lehman L, Silva I, Li Q, Johnson A E, Mark R G. AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. *Computing in Cardiology*. 2017; 44: in press.
- [4] Li Q, Liu C Y, Oster J, Clifford G D 2016 *Machine Learning for Healthcare Technologies*, ed D A Clifton (London: IET) pp 33-58
- [5] Oster J, Clifford G D. Impact of the presence of noise on RR interval-based atrial fibrillation detection. *J Electrocardiol*. 2015; 48: 947-951.
- [6] Colloca R, Johnson A E W, Mainardi L, Clifford G D. A support vector machine approach for reliable detection of atrial fibrillation events. *Computing in Cardiology*. 2013; 40: 1047-1050.
- [7] Walton S, Hassan O, Morgan K, Brown M R. Modified cuckoo search: A new gradient free optimisation algorithm. *Chaos, Solitons & Fractals*. 2011; 44: 710-718.
- [8] Whitaker B W, Suresha P B, Liu C Y, Clifford G D, Anderson D V. Combining sparse coding and time-domain features for heart sound classification. *Physiological Measurement*. 2017; 38: 1701-1729.
- [9] Clifford G D, Lopez D, Li Q, Rezek I. Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments. *Computing in Cardiology*. 2011; 38: 285-288.
- [10] Li Q, Mark R G, Clifford G D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol Meas*. 2008; 29: 15-32.
- [11] Linker D T. Long-term monitoring for detection of atrial fibrillation. US Patent. 2009; US20060084883:
- [12] Sarkar S, Ritscher D, Mehra R. A detector for a chronic implantable atrial tachyarrhythmia monitor. *IEEE Trans Biomed Eng*. 2008; 55: 1219-1224.
- [13] Liu C Y, Oster J, Reinertsen E, Li Q, Zhao L N, Nemati S, Clifford G D. A review of methods of entropy for atrial brillation detection. *BioMedical Engineering OnLine*. 2017; submitted.
- [14] Chang C C, Lin C J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011; 2: 1-27.

Address for correspondence.

Chengyu Liu
School of Instrument Science and Engineering
Southeast University
Nanjing, China
E-mail: bestlyc@sdu.edu.cn