Cardiac Rhythm Classification from a Short Single Lead ECG Recording via Random Forest

Ruhi Mahajan[1], Rishikesan Kamaleswaran[1], John Andrew Howe[2], Oguz Akbilgic[1]

[1]University of Tennessee Health Science Center, Memphis, TN, USA
[2]Independent Researcher, Riyadh, Saudi Arabia

Abstract

Detection of atrial fibrillation (AF) from electrocardiogram (ECG) recordings is one of the prevailing challenges in the field of cardiac computing. The task of the PhysioNet/Computing in Cardiology 2017 challenge is to distinguish the AF rhythms from non-AF rhythms using a short single lead ECG recording. In this study, we analyzed 62 time and frequency-domain, linear, and nonlinear features to discriminate four classes, viz., normal sinus rhythm, AF, noisy, or other rhythm. The feature space dimension was reduced to 37 using a Genetic Algorithm based feature selection. We trained a random forest classifier on the given 8,528 training dataset and obtained a ten-fold cross validation classification accuracy of 82.7%. On the test dataset, we obtained an F_1-score of 0.91, 0.74, and 0.70 for NSR, AF, and other rhythms, respectively. Results suggest that with the proposed model it is possible to classify cardiac abnormalities from a single lead ECG even when the recordings are of short duration.

1. Introduction

Arrhythmias are described as abnormalities of the heart, such as slow, fast, or irregular rhythm. Arrhythmias are caused by problems with the electrical activities in the heart that typically maintain a steady heartbeat. Although there are several types of arrhythmias, AF is the most common type of serious arrhythmia, affecting millions of people worldwide [1-3]. Identification of the type of arrhythmia is necessary for an effective treatment.

The electrocardiogram (ECG) is the most commonly used diagnostic tool in identifying abnormalities in heart rhythms. Typically, an expert cardiologist can identify abnormal heart activities in ECG. However, in some cases, distinguishing among different types of arrhythmia is challenging, especially with continuous data collection and in the presence of signal noise. To mitigate the cognitive challenges of computing multiple aspects of the ECG, statistics and machine learning based decision support tools can assist cardiologist in their decision-making.

In this paper, we developed a model using features extracted from various techniques including Sample Entropy (SampEn), Probabilistic Symbolic Pattern Recognition (PSPR), and multiwavelet decomposition to classify the ECGs as 1) normal sinus rhythm (NSR), 2) AF, 3) Other rhythm or 4) Too noisy to classify. All non-AF abnormal rhythms were considered in the Other rhythm category. We propose to use a combination of linear, nonlinear, time and frequency domain features to boost the classification performance of ECG rhythm detection.

The paper is organized as follows. Section 2 provides an insight of the dataset and pre-processing steps used in the analysis. In Section 3, we describe the feature extraction and feature selection techniques used along with the classifier. The key results are discussed in Section 4, followed by the concluding remarks.

2. Material and Methods

This section describes the dataset used in this study and gives the background of some of the concepts used in the classification algorithm.

2.1. Dataset

A total of 8,528 single lead ECG recordings were provided in the training dataset of the PhysioNet/ Computing in Cardiology Challenge 2017 [4-5]. ECG recordings, collected using an AliveCor device, were sampled at 300 Hz and had a minimum, maximum, and mean duration of 9.0 s, 61.0 s, and 32.5 s, respectively.

In the training dataset, there were 5,050 NSR, 738 AF, 2,456 Other, and 284 Noisy ECG recordings. A hidden test dataset containing 3,658 ECG recordings was used to evaluate the performance of the proposed classification model. Labels for each class were provided by AliveCor and later revised by the challenge organizers.

2.2. Pre-processing of ECG

All ECG recordings were band-pass filtered between 5-26 Hz to remove baseline wandering, power-line

 ISSN: 2325-887X DOI:10.22489/CinC.2017.179-403

interference, and to maximize the QRS energy. ECG recordings were also down-sampled to 200 Hz to decrease the computation time in feature extraction. We developed a protocol to statistically identify and exclude any abnormal spikes in the recordings caused by electrode disconnect, physiological artifacts, etc. We also developed a method in MATLAB to identify and correct inverted QRS complexes in ECG recordings.

2.3. QRS complex detection

The beat-to-beat intervals (RR) extracted from the QRS complex provide valuable information on the cardiac-autonomic function in healthy and disease states. We used the Pan-Tompkins [6] algorithm to extract RR intervals from the ECG recordings sampled at 200 Hz. The algorithm works well for a clean ECG; however, the algorithm misses many R peaks with the presence of noise and inverted QRS. Therefore, we first identified and corrected the inverted QRS complexes and then ran the Pan-Tompkins algorithm yielding drastically improved accuracy in QRS detection (refer Fig. 1).

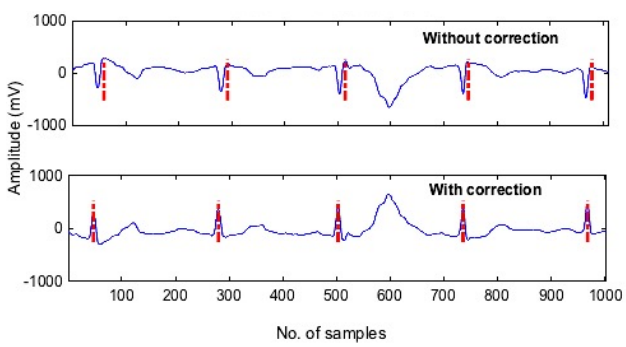

Figure 1. QRS complexes identified on a representative ECG recording without and with lead reversal correction.

3. Procedure

In this section, we discuss the process of feature extraction from ECG recordings. We also briefly explain our feature selection technique based on a genetic algorithm (GA), and the classifier used for distinguishing different rhythms.

3.1. Feature extraction

To effectively extract the underlying information present in ECG recordings, we explored possible linear measures which are robust to noise. We implemented PSPR, a linear feature extraction approach, to discretize and model symbolic pattern transitions in ECG recordings [7-9]. We used five symbols {a, b, c, d, e} to discretize ECG; then we used probability theory to learn dynamics in the ECG morphology. Our previous study shows that PSPR performs even better if ECG recordings are sampled at a low frequency [10]. Therefore, in this study, we down-sampled ECG recordings at 8 Hz to extract the PSPR features. Four PSPR features were calculated by comparing the similarity of the discretized series with a reference AF episode database of 25 ECG recordings [11] and 100 NSR ECG recordings of the training dataset.

To gain more information from the nonstationary ECG recordings, we extracted five features with wavelet decomposition. Based on our preliminary observations, we decided to apply the *Symlet* wavelet and decomposed the recordings up to level 5. Features obtained from the variance of three detail coefficients (cd_3, cd_4, cd_5) and variance of the autocorrelation of (cd_1, cd_2) were found to provide significant discriminative information and hence were used in the feature space.

A nonlinear metric like SampEn conveys information pertaining to predictability and chaotic behavior of a signal [12-13]. Therefore, to gauge the complex dynamics of ECG recordings, we computed SampEn values. Because ECG signals are intrinsically non-stationary, features extracted from the entire series may be locally unsuitable. Hence, we segmented each ECG recording of length *L* into epochs of one second and computed SampEn for each epoch sequentially using equation 1. This resulted in SampEn series (each of length L-200) for the corresponding ECG series.

$$\text{SampEn}(\text{m}, \text{r}, \text{N}) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right]. \tag{1}$$

where, *m*=3, is the maximum epoch length to be compared, *r*=0.25, is the tolerance window and N=200, is the length of the ECG series. $B^m(r)$ is the probability that two sequences will match for *m* points, whereas $A^m(r)$ is the probability that the two sequences will match for *m+1* points. The mean and range of SampEn were used as additional features.

Furthermore, we evaluated frequency-domain features from spectral analysis of ECG recordings. We computed the average power in very low frequency (VLF) = 0-0.04 Hz, low frequency (LF)= 0.04-0.15 Hz, high frequency (HF)= 0.15-0.5 Hz, and a ratio of LF to HF. Among these features, VLF provided the strongest discriminative power.

In addition to the abovementioned features, we also evaluated conventional time-domain measures such as pNN50 (the percentage of beats having a difference greater than 50ms); descriptive metrics such as mean, median, kurtosis, standard deviation, range and skewness of preprocessed ECG recordings and original, 1st order, and 2nd order RR interval series. We also used the averages of RR intervals falling in the lower 5% and higher 5% tails of their frequency distribution. We extracted another set of temporal features by performing the Kolmogorov–Smirnov (KS) test to compare the given RR interval with the reference RR interval database created from 50 examples from each class. We then computed features using the mean *p-value* comparing AF and Other.

In total, we identified 62 different features to fetch morphological, temporal, and spectral information from ECG recordings. The statistical metrics used are very simple, yet they provide vital information for ECG rhythm classification. The proposed features in this study are computationally inexpensive and therefore, can also be utilized for real-time arrhythmia detection.

3.2. GA-based Feature selection

To enhance the generalizability of the classification model, we removed redundant features using a GA-based feature dimension reduction approach. The GA is an optimization technique to obtain the subset of features that maximizes the predictive accuracy. Our algorithm was based on Babtunde's model [14]. A population of 10 individuals was created and the initial population was randomly generated. For every individual in the current population, an ensemble of bagged decision trees with 220 learners was estimated. The resulting random forest was then tested with *ten*-fold cross validation. The fitness function of the individual was based on the F_1-score for NSR, AF, and Other rhythms. Once the fitness of all individuals of the current generation was computed, the GA generated the next generation. This process was iteratively executed for 15 generations.

3.3. Random Forest classifier

We implemented a random forest classifier to discriminate four classes of ECG rhythms. Individual decision trees tend to overfit, so we used bootstrap-aggregated (bagged) decision trees, aggregating the predictions from an ensemble of 220 decision trees. The random forest classifier selects a random subset of features to use at each decision split, which helps reduce the correlation between decision trees. Further, to avoid overfitting of the classification model, we performed *ten*-fold cross-validation in which we trained our model on nine equal-size subsamples of the training data and tested it on one subsample. The reported classification statistics in this paper are the average of *ten*-fold cross validations. We also compared the classification results with other classifiers (Support vector machine, discriminant analysis, decision trees), but the random forest classifier provided the best classification accuracy by far.

4. Results

This paper addresses a challenging problem of reliably detecting abnormal cardiac rhythms from the broad taxonomy of rhythms. The training dataset provided contained single lead, noisy ECG recordings of very short duration. We evaluated different features to characterize the behavior of ECG rhythms in both time and frequency domains. The hybrid scheme of time, frequency, linear, and nonlinear feature extraction techniques improved the classification accuracy of ECG recordings contaminated with artifacts and noise.

Our GA-based algorithm selected 37 discernable features from the pool of 62 features. The final feature matrix with 37 features was used as input to the random forest classifier to yield a classification output as: Y=0 (NSR), Y=1 (AF), Y= 2 (Other), and Y= 4 (Noisy) rhythms. Table I depicts the list of features selected in the final classification model. For illustration purposes, box plots of two features- pNN50 and range of sample entropy computed from the training dataset are shown in Fig. 2. As depicted in the plot, the mean percentage of beats differing by more than 50 ms is low for the case of normal rhythm. Also, the mean range of sample entropy is relatively high for the noisy rhythms, which suggests the chaotic behavior in noisy ECG recordings.

The objective of this paper was to emphasize the correct classification of usable recordings. So, to reduce the effective weight assigned to noisy recordings, F_1-scores were calculated from only the three main classes and the noisy class was ignored. The overall F_1 score, therefore, was computed as:

$$\text{Overall } F_1\text{-score} = \frac{F_{1n}+F_{1a}+F_{1o}}{3} \tag{2}$$

where $F_{1n} = \frac{2*Precision*Recall}{Precision+Recall}$ is the F_1 score of the normal rhythm. Likewise, F_{1a} and F_{1o} were calculated for AF and Other rhythms, respectively. F_1 scores obtained from the training and hidden test dataset are recorded in Table 2. The results suggest that the classification model doesn't overfit and is generalizable.

Table 1. Summary of final selected features for discriminating four ECG rhythms in the model.

Features used in this study
Time-domain
Descriptive measures of ECG
Descriptive measures of RR intervals
Descriptive measures of 1st order RR intervals
Descriptive measures of 2nd order RR intervals
Quantile based on RR intervals
KS test based on RR intervals
Heart rate variability measures
Frequency-domain
VLF power
Linear
Wavelet coefficients
PSPR
Nonlinear
SampEn from one second epoch of ECG

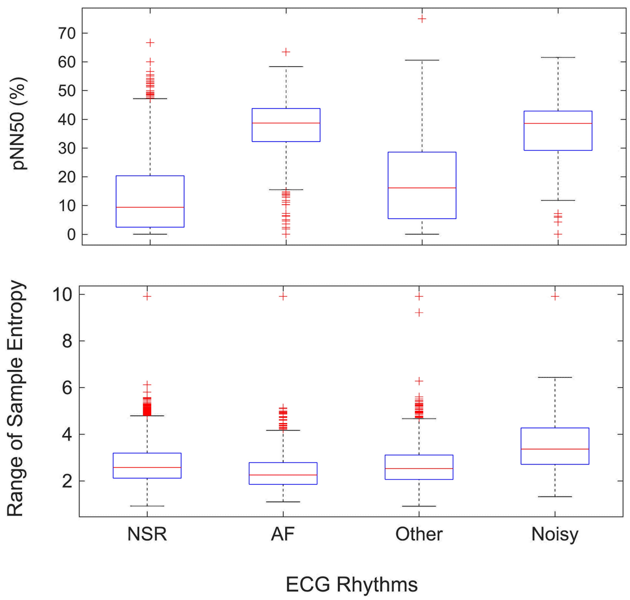

Figure 2. Box plot of two representative features used for discriminating the four classes of rhythms.

Table 2. F_1-scores obtained with the proposed classification model.

Rhythm Type	F_1- Score	
	Training dataset	Test dataset
Normal	0.89	0.91
Atrial fibrillation	0.76	0.74
Other	0.72	0.70
Average	*0.79*	*0.78*

5. Conclusions

In this study, we extracted and analyzed various morphological, time and frequency-domain features to characterize changes in ECG recordings. We used the PhysioNet 2017 challenge dataset, which contained 8,528 single lead ECG recordings lasting from 9 to over 60 seconds. We implemented a GA-based feature selection technique which resulted in a reduced feature space of 37 features to classify different ECG rhythms. Using a random forest classifier, we obtained an overall F_1-score (for NSR, AF and Other rhythms) of 0.79 on the given training dataset and 0.78 on the hidden test dataset. Our results demonstrate that the efficacy of AF detection can be enhanced by using the proposed machine learning approach.

References

[1] A. J. Camm et al., "Guidelines for the management of atrial fibrillation," *Eur. Heart J.*, vol. 31, no. 19, pp. 2369–2429, 2010.

[2] W. . Kannel, P. . Wolf, E. . Benjamin, and D. Levy, "Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates," *Am. J. Cardiol.*, vol. 82, no. 7, p. 2N–9N, 1998.

[3] G. V Naccarelli, H. Varker, J. Lin, and K. L. Schulman, "Increasing Prevalence of Atrial Fibrillation and Flutter in the United States," *Am. J. Cardiol.*, vol. 104, no. 11, pp. 1534–1539, 2009.

[4] G. Clifford et al.,"AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017," Computing in Cardiology, vol. 44, 2017 (In Press).

[5] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. E215-20, 2000.

[6] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985.

[7] O. Akbilgic, A. Howe, and R. Davis, "Clustering Atrial Fibrillation via Symbolic Pattern Recognition," *J. Med. Stat. Informatics*, vol. 4, pp. 1–9, 2016.

[8] O. Akbilgic and A. Howe, "Symbolic Pattern Recognition for Sequential Data," *Seq. Anal.*, vol. In Press, 2017.

[9] R.Mahajan, T. Viangteeravat, and O. Akbilgic. "Improved Detection of Congestive Heart Failure via Probabilistic Symbolic Pattern Recognition and Heart Rate Variability Metrics." *International J. of Medical Informatics,* 2017, vol. 108, pp. 55-63.

[10] R. Mahajan, R. Kamaleswaran, and O. Akbilgic, "Effects of Varying Sampling Frequency on the Analysis of Continuous ECG Data Streams," in *Data Management and Analytics for Medicine and Healthcare*, 2017, pp. 73–87.

[11] G. B. Moody, A. L. Goldberger, S. McClennen, and S. P. Swiryn, "Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001," *Comput. Cardiol.*, pp. 113–116, 2001.

[12] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, pp. R789-97, 2002.

[13] R. Alcaraz and J. J. Rieta, "A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms," *Biomedical Signal Processing and Control*, vol. 5, no. 1. pp. 1–14, 2010.

[14] O. Babatunde et al, "A genetic algorithm-based feature selection," *Br. J. Math. Comput. Sci.*, vol. 4, no. 21, pp. 889–905, 2014.

Address for correspondence:

Oguz Akbilgic
LeBonheur Research Center, 50 N Dunlap Ave, Suite 490 R
Memphis, TN, 38103, USA
oakbilg1@uthsc.edu