IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON TRENDS, PERSPECTIVES AND PROSPECTS OF MACHINE LEARNING APPLIED TO BIOMEDICAL SYSTEMS IN INTERNET OF MEDICAL THINGS

Received July 22, 2018, accepted September 8, 2018, date of publication September 19, 2018, date of current version October 17, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871220

Automatic Atrial Fibrillation Detection Based on Heart Rate Variability and Spectral Features

ZHENNING MEI[1], (Student Member, IEEE), XIAO GU[1], (Student Member, IEEE), HONGYU CHEN[2], AND WEI CHEN[1,3], (Senior Member, IEEE)

[1]Center for Intelligent Medical Electronics, School of Information Science and Engineering, Fudan University, Shanghai 200433, China
[2]Department of Industrial Design, Technical University of Eindhoven, 5600 MB Eindhoven, The Netherlands
[3]Shanghai Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention, Fudan University, Shanghai 200032, China

Corresponding author: Wei Chen (w_chen@fudan.edu.cn)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFE0112000 and in part by the Shanghai Municipal Science and Technology Major Project under Grant 2017SHZDZX01.

ABSTRACT Atrial fibrillation (AF) is one of the most common sustained arrhythmias, affecting about 1% of the population around the world. Rapid popularization of portable and wearable devices in recent years makes widespread personalized and mobile healthcare get closer to reality than ever before. This paper presents a method aiming for automatic detection of AF from short single lead electrocardiogram (ECG) recordings. Since AF is a kind of arrhythmia being likely to alter the dynamics of heart rhythms and/or the morphological characteristics in ECG tracings, heart rate variability (HRV)-based metrics and frequency analysis are adopted as feature extractors. We validate our method on a public available data set comprised of short ECG recordings of normal rhythm (N), AF (A), and other arrhythmias (O) by support vector machine and bagging trees. For two-class classification problems (N versus A), accuracy varies from 92.0% to 96.6% under different additional noise levels. For three-class classification problem (N versus A versus O), accuracy as high as 82.0% is obtained. Experimental results suggest than even for a relatively short ECG recording, nonlinear descriptors of HRV are still efficient and robust for AF detection.

INDEX TERMS Atrial fibrillation, ECG, heart rate variability, biomedical signal processing, machine learning.

I. INTRODUCTION

Atrial fibrillation (AF) is one of the most common sustained arrhythmias in the world with the elderly being more likely to be affected. It has been reported that more than 5% of people older than 65 suffer from this disorder. Men are more vulnerable to AF than women [1]. AF may lead to stoke and heart failure, so timely diagnosis and control are crucial in the management of AF.

Atrial fibrillations can be classified into paroxysmal, persistent and permanent according to the medical history of patients and whether it can be terminated spontaneously or by external interventions. Although an abnormal tissue as substrate is identified as an important factor maintaining the arrhythmia, AF is also observed among patients without any structural heart disease [2]. After persistent atrial fibrillation, the atrium of patients could adapt to this pathological condition in the aspects of ion channel functions of cardiomyocyte, called remodeling. The shortened refractory period (RP) of the action potential (AP) is a typical electrophysiological alteration caused by AF leads to tachycardia and possibly the absence of P waves, which can be reflected through electrocardiogram (ECG) tracing. And thus ECG is a widely used tool for the diagnosis of AF.

The ultimate goal of the treatment of AF is the restoration of sinus rhythm. Pharmacological cardioversions and external/internal electrical cardioversions are efficient for recent onset of AF and persistent AF respectively. For AF initiated by some foci, for example, pulmonary vein, known as being responsible for the onset of left atrial tachycardia, radiofrequency ablation causes the necrosis of specific region so the tissue in this region will not be electrically active again. In addition, surgery and implantation of pacemaker are also options [2].

As for the diagnosis of AF, detection of abnormal episodes in ECG recordings provides useful information to clinicians and is usually served as the first step of the diagnosis procedure. In clinical practice, experienced cardiologists inspect ECG tracing manually, which is a time-consuming work.

For long-term recording acquired by holter monitors, it is especially hard for cardiologists to locate the abnormal episodes. So it is an urgent need to develop efficient and robust decision support systems for automatic detection of atrial fibrillation (ADAF). Most of these kinds of decision support systems take ECG signals as inputs. Then an algorithm is performed to translate the raw data into useful features and finally a graspable result for clinicians is outputted. Different kinds of features inspired by medical knowledge or driven by data analytics have been adopted in related research [3]. Among them, the abnormal P wave and QRS morphology are widely validated by clinicians as indicative signs of AF with physiological interpretability [4]. Instead of encoding the medical knowledge explicitly, features of a data-driven nature are also adopted. For example, inter-pulse intervals (R-R intervals, RRI), defined as the durations between consecutive R peaks, code information about not only coarse-grained heart rate (HR) as well as dynamics characteristics of heart activities. RRI based features, including heart rate variability (HRV) [5], are investigated in detail [6] for ADAF. Furthermore, frequency spectrum analysis and nonlinear dynamics analysis about both raw data and some discrete time series like RR intervals are also proposed to be effective feature extraction methods [7], [8]. These features are usually not intuitionally like the morphology of characteristic waves (CW) in ECG. Quantitative metrics are indispensable before such kind of information entering into medical context. All the automatic detection methods mentioned above are based on the recording and analysis of ECG signal. Holter monitoring and implantable pacemaker capable of recording and storing the ECG signals have been adopted for many years. However, the application of such technologies is still limited because the normal use of them is, to some extent, obtrusive. The prevalence of wearable and portable devices aiming for continuous and unobtrusive bio-signal monitoring and health supports has been witnessed during recent years. For example, many wearable and portable ECG monitoring systems have been proposed [9]–[15]. The interfaces between the system and human body are diversified. Most of the interfaces are of contact-type. Except for pre-gelled Ag/AgCl electrode, other kind of electrodes made of textile [9], [12], [14], polymer materials [11], are reported. Non-contact ECG sensing, mainly based on capacitive coupling, are also studied [13]. The skin-to-electrode interface can make significant impact on the quality of acquired signal [15]. Camera based and bio-radar based modalities have also been investigated to provide alternatives to reflect related information about heart dynamics.

For state-of-the-art wearable and portable devices working under highly-varied unstructured environments, a relatively low signal-to-noise ratio (SNR) of the observations is, if not inevitable, frequently encountered. A robust AAFD algorithm is expected to be able to extract reliable information from ECG signal of heterogeneous quality. Some characteristics waves (CW) extracted with the aid of standard multi-lead ECG monitor are regarded as informative features and are deemed to have medical implications may be unavailable, or at least unstable in unstructured environments. Instead, only single lead ECG recording, often interrupted by motion artefacts is desirable. From the above, this paper presents an AAFD method suitable for ubiquitous computing of data collected from wearable devices in non-structured environments. Heart rate variability based metrics and frequency features are adopted to reflect the abnormal heart rhythm and morphological change during AF respectively. Supervised machine learning models are trained with the aid of a public dataset comprised of short single lead ECG recordings collected by AliveCore [16], [17], a kind of portable ECG monitoring device. The sampling frequency is 300 Hz and build-in bandpass filter is adopted. The remainder is organized as follows. Section II provides details about feature extraction and dataset. The results under different experimental settings are given in Section III. Section IV is discussion part investigate the advantages of the proposed method and future work, followed by Section V as conclusion.

II. MATERIALS AND METHODS

Our aim is to develop a robust AAFD method processing data collected from wearable or portable devices in non-structured environments. Data collected from such devices are usually vulnerable to different kinds of artifacts. A relatively long and clean section of signal cannot always be anticipated from wearable systems because serious artifacts could scatter among continuous recording and thus interrupt available signal into smaller slices. Portable devices, on the other hand, do not always work in a wearable fashion or attached to the patients. The length of recording depends on the users' behavior so is unmanageable for the developers. Therefore, morphological characteristics including QRS complex length, PR/ST intervals etc. even being clinically relevant and physiologically sound, are not preferred because these subtle features cannot always be guaranteed in the scenarios mentioned above.

However, R peak, the most dominant fiducial point in ECG signal, is usually the most unaffected feature in ECG. If the R peaks are not identifiable in a segment of ECG, the other fiducial points or CWs are more unlikely to provide useful information then. On the contrary, when other fiducial points or CWs are not available, R peak is still possible to be detected with high confidence and accuracy, typically happening under low or moderate noise level. Therefore, heart rate variability (HRV) based features are considered for their robustness under noisy scenarios. In addition, HRV based features code only dynamics characteristics about heart activities. Detailed information carried by ECG signals, although vulnerable in time domain, can be preserved partially in frequency domain. In this, spectral features are adopted as supplementary. Figure 1 pictures the overall framework of the proposed method.

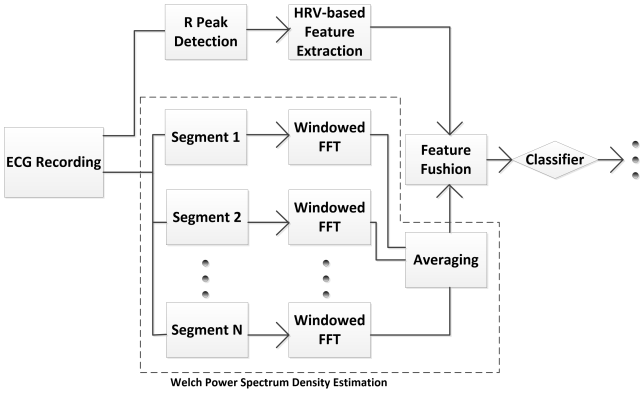

FIGURE 1. The framework of the proposed method.

A. DATA DESCRIPTION

In this work, a public available dataset released as the material for Computing in Cardiology Challenge (CinC) 2017 is used [18]. The public part consists of 8528 single lead ECG recordings with length varying from 9 seconds to over 1 minute. The labeling of the recordings is performed by 4 clinical cardiologists independently. About 60% of the data are classified as normal rhythm (N), 9% of the data are diagnosed as episodes of atrial fibrillation (A). About 28% of the data belong to arrhythmias other than AF (O), with no more detailed information provided. Finally, about 3% of the data are identified as too noisy to be useful (∼). The distribution of the dataset and sample recordings are depicted in Table 1 and Figure 2.

TABLE 1. Class distribution of the dataset.

Class	Numbers of recordings	Portion
Normal Rhythm(N)	5076	59.52%
Other Arrhythmias(O)	2415	28.32%
Atrial Fibrillation (A)	758	8.89%
Noisy (∼)	279	3.27%

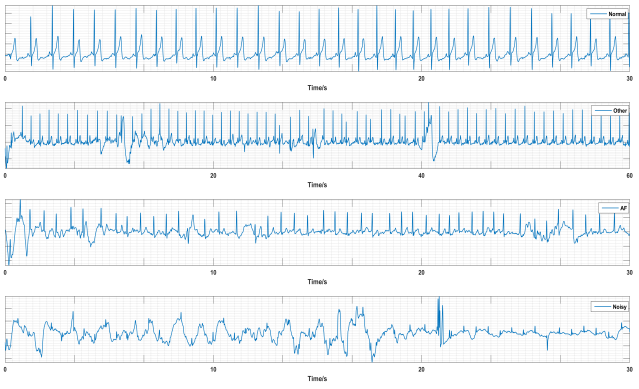

FIGURE 2. Sample recordings from each class.

B. FEATURE EXTRACTION

1) HRV BASED FEATURES

R peak detection is the first step to extract HRV based features. A famous real-time QRS detection algorithm developed by Pan and Tompkins (P&T) is adopted for its superior performance [19]. The indices of R peaks in ECG recording are identified. Denote the number of R peaks detected as N, the sampling frequency of the signal as f Hz and the length of this recording as L. The average HR (beats/min) is obtained as $HR = 60 \cdot \frac{N \cdot f}{L}$. And a sequence RRI comprised of RRI in order:$RRI = [RR_1, RR_2, \ldots, RR_{N-1}]$.

Heart rate variability (HRV) is the oscillations between consecutive heart beat intervals. HRV was not regarded as a digital signal, for all intents and purposes, when it was observed and conceptualized in medical context [20] for the first time. However, frequency analysis was introduced later for the relation between the activities of autonomous nervous system (ANS) and the energy distribution of HRV in frequency domain. HRV is believed to be modulated by both sympathetic and vagal nervous systems [5]. So the power spectrum density (PSD) of HRV can be interpreted in terms of sympathetic-vagal balance. However there are some intrinsic flaws for HRV being evaluated as a digital signal sampled in some manner. A typical RRI sequence as a function about time and the original ECG from which the RRI sequence is extracted are provided in Figure 3. Only after the next R peak is identified, the last RRI can thus be determined, and if the human hearts beat in a strict steady rhythm, there will be no HRV. The RRI sequence, if being taken as a digital signal, is irregularly spaced in time with dimension of quantity being also the time instead of mass of length etc. Interpolation and resampling [21] or Lomb-Scargle Periodogram [22] are frequently used techniques to circumvent this unevenness in time domain. However, these methods assume the existence of a *hypothetical* continuous random process from which any evenly or unevenly spaced signal can be sampled depending on our choice. Therefore, subtle side effects are introduced by different interpolation criteria. Of course, the evenly spaced fashion is welcomed under almost all circumstances and no additional operations are necessary but it should be noticed that the existence of such a continous random process is, although intuitive in many fields including but not limited to astro-observations [22], not self-evident for HRV. When we want to take it as a function about time, it becomes a peculiar quantity quantified in the unit of time while also varying as time elapses. A better model could be point process, which circumvent the explict or implict interpolation operations among existing methods in form by providing parametric probability distributions whereby quantities of interest can be calculated analytically from the formula with estimated coefficients. However, the advantages of point process based heart rate variability analysis is still formalistic because no brand new metrics has been derived from it and the unsatisfactory interpolation was just replaced with a more subtle assumption instead of truly removed.

Except for the conceptual difficulties, an acceptable frequency resolution could not always be satisfied in wearable scenarios. Typical value of length of recording is significantly lower than that was recommended by EHA (≥ 5 *minutes*) [5]. Meanwhile, the counterpart of the famous *uncertainty*

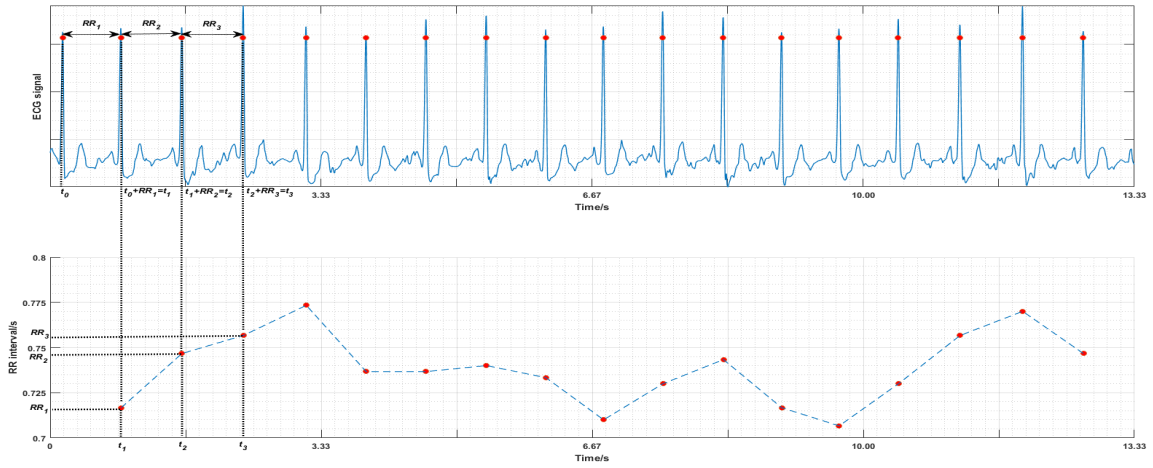

FIGURE 3. HRV from ECG: (a) Original signal. (b) Inter pulse intervals as an unevenly spaced signal.

principle of quantum physics in signal processing [23] asserts the time resolution and frequency resolution of a given signal cannot be arbitrarily high concurrently. Given a signal $s(t)$ with normalized energy distribution on time axis ($\int_{-\infty}^{\infty} |s(t)|^2 dt = 1$), we have:

$$\left(\int_R t^2 |s(t)|^2 dt\right) \cdot \left(\int_R f^2 |\hat{s}(f)|^2 df\right) \leq \frac{1}{16\pi^2} \tag{1}$$

where $\hat{s}(f)$ is the *Fourier transform* of $s(t)$. For signals with very compact supports in time domain like RRI sequences obtained here, the energy concentrations of the corresponding signal in frequency domain is poor.

According to the reasons mentioned above, direct power spectrum density estimation is waived in this research. Instead, *RRI* sequence and its first order difference are arranged into patternized matrix and mapped into 2-D plane (Poincare Plot) and then several descriptors about the patternized matrix and Poincare Plot are supposed to be able to characterize the detailed dynamical information carried.

Denote the first order difference of *RRI* as *dRR* where $dRR_i = RR_{i+1} - RR_i$.

A -1 Hankel matrix of order n is a matrix defined by its first row or column. Given a series $A = [a_1, a_2, a_3, \ldots, a_n]$, the -1 Hankel matrix defined by A is:

$$H = \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_2 & a_3 & a_4 & \ldots & a_1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{n-1} & a_n & a_1 & \ldots & a_{n-2} \\ a_n & a_1 & a_2 & \ldots & a_{n-1} \end{bmatrix} \tag{2}$$

Such a special Hankel matrix is *Hermitian symmetry* so it can be diagonalized by a *unitary* matrix U:

$$H = U^* D U \tag{3}$$

And the $*$ denotes the conjugate transpose and D is the diagonal matrix with the eigenvalues of H arranged on its diagonal line. We found that the spectrum of H is also central

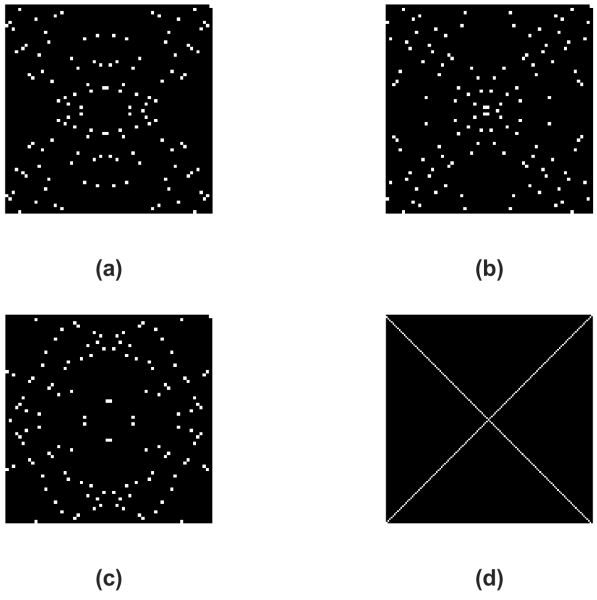

FIGURE 4. Sample HankDist for: (a) Normal (Record A00003), (b) AF (Record A00005), (c) Other arrhythmia (Record A00008), (d) Sinusoidal signal.

symmetry and the eigenvectors constituting U are bases in *Fourier* domain. Therefore, we arrange the eigenvalues by ascending order and permute the rows and columns in U^* and U correspondingly. A Discrete Fourier Transform (DFT) is then performed on U and a feature map is obtained (see Figure 4, note that only two non-zeros entries exist for the DFT of each column).

Since the feature map is symmetry, we proposed a descriptor [24] to characterize the structural irregularity of the original time series A as follows, denote the row index of the non-zero entry in the i_{th} column of U as $I_i, i \in 1, 2, \ldots, \frac{n}{2}$; the proposed descriptor *HankDist* is defined as:

$$\textbf{HankDist} = \frac{4}{n^2} \sum_{i=1}^{[\frac{n}{2}]} |I_i - i| \tag{4}$$

The non-zero entries in the feature maps of pure sinusoidal series lie in the diagonal line and anti diagonal line, so the *HankDist* for such series will be zero. However, for white noise, the non-zeros entries in the feature map scattered irregularly and the *HankDist* is large. The *HankDist* is computed for both *RRI* and *dRR*.

Among the standard descriptors [5] about Poincare plot SD_1, SD_2 and $SD_{ratio} = \frac{SD_1}{SD_2}$ are linear descriptor of the *RRI* sequence. Given consecutive *RRI* pairs (RR_i, RR_{i+1}), RR_{i+1} are plotted against RR_i. In a Poincare plot constructed from normal *RRI*, most of the data points should distribute in the neighbor of the straight line $y = x$ and the data cloud should form an ellipse. Usually these data points are used to fit the ellipse and the long axis (SD_1) and short axis (SD_2) are believed to reflect the long-term and short-term dispersion of the fluctuation of *RRI* so also the dynamics of heart activity. Since the length of *RRI* sequence is very limited, an *embedding* step is skipped here. Denote $(x, y)^T = (RR_i, RR_{i+1})^T, i = 1, 2, \ldots, N-1$, the data points are rotated 45 degree in clockwise first to extract SD_1 and SD_2:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ -\sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{5}$$

where the $(x', y')^T$ denotes the coordinates of new data points. And the variance along x' axes are computed as SD_1, and $SD_2^2 = 2 \cdot Var(RR) - SD_1^2$ [25].

It has been discovered that although Poincare plot is believed to be of nonlinear analysis tool. Such descriptors (SD_1, SD_2) are, in fact, *linear combination* of the basic statistics of the original *RRI* sequence [25]. Since they can be fully determined by the second order statistics of the original *RRI* series, they are not *nonlinear* in their nature. Furthermore, it is obvious that such descriptors are invariant under some permutations of original *RRI* series, (e.g. when $RR_1 = RR_{N-1}$, any shift of original *RRI* yield exactly the same data cloud on 2-D plane), implying limited ability about reflecting the order structure in *RRI*.

Complex correlation measure (CCM) is proposed for encoding temporal information of RR sequence and as a real nonlinear descriptor [26]. CCM traces the 'generation' of the Poincare plot by calculating the oriented area spanned by consecutive 3 *RRI* pairs (after possible delay-embedding). Denote $A(i)$ as the oriented area spanned by $((RR_i, RR_{i+1}) \rightarrow (RR_{i+1}, RR_{i+2}) \rightarrow (RR_{i+2}, RR_{i+3}))$, we have:

$$A(i) = \begin{vmatrix} RR_i & RR_{i+1} & 1 \\ RR_{i+1} & RR_{i+2} & 1 \\ RR_{i+2} & RR_{i+3} & 1 \end{vmatrix} \tag{6}$$

And CCM is the normalized oriented area:

$$CCM = \frac{1}{K\pi SD_1 \cdot SD_2} \cdot \sum_{i=1}^{K} A(i) \tag{7}$$

Different from *RRI* for which a characteristic power spectrum distribution with interpretable frequency components is expectable [5], the PSD of the difference of RRI sequences (δRR) reveals no significant distinction with respect to that of white noise. However, δRR was found being capable of distinguish AF from normal rhythm [26]. AFEv is extracted from the Poincare plot of the δRR and has been validated as a powerful AF detector on baseline datasets including MIT-BIH AF and MIT-BIH NSR [27]. Finally, RR mean and RR std are also calculated.

2) SPECTRAL FEATURE

The HRV based features contain only sketchy information about heart activities. Although the detailed morphology of ECG is vulnerable to noise so it is not a good idea to characterize the CW in time domain, the structure is relatively stable in frequency domain. Welch power spectrum density (PSD) estimation is a modification of periodogram method [28]. Compared with naïve periodogram method, reduced variance of the estimation is obtained at the expense of the reduction of frequency resolution.

Suppose we have a uniformly sampled ECG signal of length N. It can be divided into M segments of length L with or without overlapping. Denote overlapping as D, we have:

$$L + (M-1)(L-D) = N \tag{8}$$

For each segment $X_k(i), i = 0, 1, \ldots, L-1$, a windowed discrete fourier transform (DFT) is calculated as follows:

$$\hat{P}_k(n) = \frac{1}{L} \sum_{i=0}^{L-1} X_k(i) W(i) e^{-2\pi j i \frac{n}{L}}, \quad k = 1, 2, \ldots, M \tag{9}$$

$W(i)$ is the window function aiming to reduce the frequency leakage. And the final spectral estimation (scaled into $[-\frac{1}{2}, \frac{1}{2}]$) is the normalized average of all the M periodograms above:

$$\hat{P}(\frac{n}{L}) = \frac{1}{M} \sum_{k=1}^{M} \frac{|L \cdot \hat{P}_k(n)|^2}{W_0}, \quad n = 1, 2, \ldots, \frac{L}{2} \tag{10}$$

where W_0 is the energy of the window function:

$$W_0 = \sum_{i=0}^{L-1} |W(i)|^2 \tag{11}$$

The frequency resolution is thus reduced from $\frac{N}{2}$ to $\frac{L}{2}$. However the variance of the estimation is also reduced by a factor up to M [28]. In our case, the order of magnitude of N is about $10^3 - 10^4$. If the order of magnitude of M is about 10^1, considering the sampling frequency is 300 Hz (so the power spectrum will distribute in 0-150 Hz), we still have a resolution level about 1 frequency point/Hz.

C. FEATURE SELECTION

Feature selection helps to choose most discriminative features from given feature set and avoid overfitting. A proper balance between the performance on training set and the scale of the whole model improves the robustness of the system and the speed of inference. Since all the features mentioned so

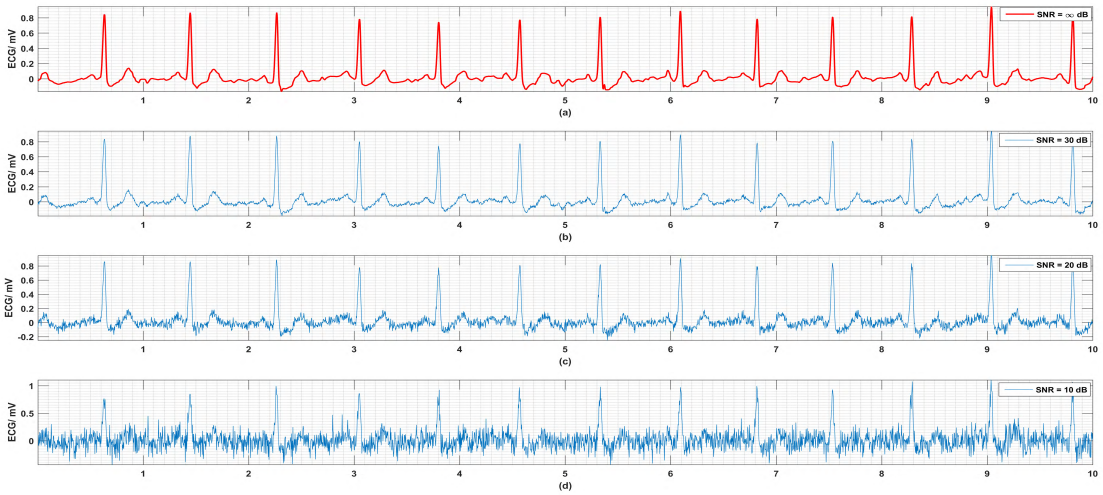

FIGURE 5. Noise test: (a) SNR = ∞ dB (original signal), (b) SNR = 30 dB, (c) SNR = 20 dB, (d) SNR = 10 dB.

far are measured at scales of intervals, a model-free feature selection scheme based on *mutual information*, called maximum-relevance-min-redundancy (MRMR) is used [29]. This scheme tries to maximize the relevance between the features and the targets (labels) while keeping the redundancies between selected features to a minimal level. Furthermore, for features being measured at the scales of nominal or ordinal, the relevance and redundancy can also be estimated with respect to mutual information.

Denotes the Ω as the set of orignal features, the S as the set of selected features, t as the targets (labels), $I(\cdot,\cdot)$ as the mutual information estimation and $|\cdot|$ the cardinality operator, the purpose of feature selection is the find the optimal subset S of cardinality m which:

$$\arg\max_{S\subseteq\Omega} I(S_m, t) \tag{12}$$

The straight forward selection procedure to find the subset S is combinatorial, which makes exhaustive search an impractical choice. Furthermore, the estimation of mutual information of high dimensional variable is difficult. In [29], incremental procedure is deployed instead as optimal first-order approximation.

Define the relevance R_V and redundancy R_D as follows:

$$R_V = \frac{1}{|S|}\sum_{i\in S} I(i,t) \tag{13}$$

$$R_D = \frac{1}{|S|^2}\sum_{i,j\in S} I(i,j) \tag{14}$$

And the object function is:

$$\arg\max_{S\subseteq\Omega,|S|=m} \left(R_V - R_D\right) \quad or \quad \frac{R_V}{R_D} \tag{15}$$

which maximizes R_V and minimizes R_D simultaneously.

D. CLASSIFICATION TASKS AND ADDITIONAL NOISE RESISTANCE TEST

Since the proportion of noisy class is very small, 279 recordings belonging to noisy class are excluded for simplicity. The remnant part is a subset consists of samples belong to one of only three different classes.

To investigate the robustness against noise of our method, noises of different intensities are artificially added. Figure 5 gives a sample recording with its noise corrupted versions. Although there might exist noises in original recordings, since we do not have a reliable estimation on the true SNR of them, we take original signals as clean signal. Then we use all the energy of original recordings to compute P_s and white noise of power P_n is added. The signal to noise ratio (SNR) is:

$$\mathbf{SNR} = 10 \cdot \log \frac{P_s}{P_n}\ \mathrm{dB} \tag{16}$$

And the whole framework depicted in Figure 1 will be performed under each noise level.

For feature values obtained under each noise level, support vector machine (SVM) and bagging trees (BT) are trained as classifiers. Since private test set is not publicly available until now, high entropic k-fold cross validation ($k = 10$) is firstly used to test the performance of classifiers on all the available data. Then 15% data of each subset (N,A,O) are reserve and only 85% of data was involved in training and validation, the performances on the remaining 15% data are reported to demonstrate the generalization ability. All the computations are performed on Matlab2015b for 10 times and averaged values are reported.

III. RESULTS

A. DESCRIPTIVE ANALYSIS

After preprocessing a 50 dimension feature vector was obtained for each recording, including the HRV based features mentioned above and the power spectrum density

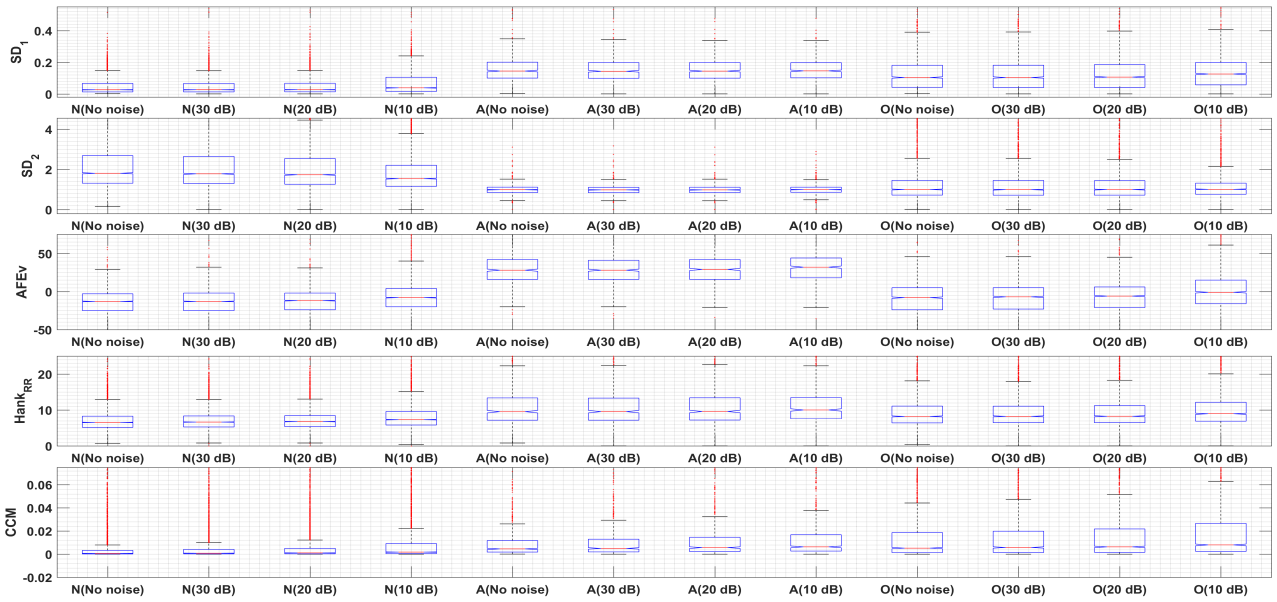

FIGURE 6. Boxplots for linear and nonlinear descriptors.

TABLE 2. Top 10 features selected by MRMR.

Rank	SNR=∞ dB	SNR = 30 dB	SNR = 20 dB	SNR = 10 dB
1	AFEv	AFEv	AFEv	RR_{mean}
2	CCM	CCM	CCM	CCM
3	SD_{ratio}	SD_{ratio}	SD_{ratio}	SD_{ratio}
4	RR_{mean}	RR_{mean}	RR_{mean}	$PSD_{4.6-6.9}$
5	$PSD_{4.6-6.9}$	$PSD_{4.6-6.9}$	$PSD_{4.6-6.9}$	$PSD_{6.9-9.2}$
6	$PSD_{6.9-9.2}$	$PSD_{6.9-9.2}$	$PSD_{6.9-9.2}$	$PSD_{9.2-11.5}$
7	$Hank_{RR}$	$Hank_{RR}$	$Hank_{RR}$	$PSD_{2.3-4.6}$
8	$PSD_{9.2-11.5}$	$PSD_{9.2-11.5}$	$PSD_{9.2-11.5}$	$PSD_{11.5-13.8}$
9	$PSD_{11.5-13.8}$	$PSD_{11.5-13.8}$	$PSD_{11.5-13.8}$	AFEv
10	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$	$PSD_{13.8-16.1}$

estimated from 0 Hz to 80 Hz. The feature selection scheme mentioned in Section II-C provided a natural ranking of the features. We finished the classification tasks defined in Section II-D using all the features, the top 10 features, the top 8 features and the top 5 features, respectively.

Table 2 provides the details about the top 10 features associated with each noise level. It was observed that the linear and nonlinear descriptors based on HRV analysis are almost the top ranked features under different noise levels. When SNR = 10 dB, the noise might disturbe the *RRI* and lead to a different ranking. Generally, HRV based features and PSD from 5 Hz to 15 Hz are those with higher relevance with class labels and bring less internal redundance than other features. Figure 6 provides the box plots about the inter-group comparison of the descriptors.

B. CLASSIFICATION PERFORMANCE

For two-class classification problems, Sensitivity, Specificity, Accuracy and AUC are well-defined and reported.

$$Sen = \frac{TP}{TP + FN} \tag{17}$$

$$Spec = \frac{TN}{TN + FP} \tag{18}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

For three-class classification problem, we defined the A as positive class and N and O as negative classes and counterparts of *Sen* and *Spec* are defined as the quotient of true positive/negative examples and all positive/negative examples and reported. Table 3 summaries the results obtained by 10-fold cross validation on the whole dataset and Table 4 provides the results obtained on artifical test set after 10-fold cross validation was performed on 85% of data during training and validation procedure.

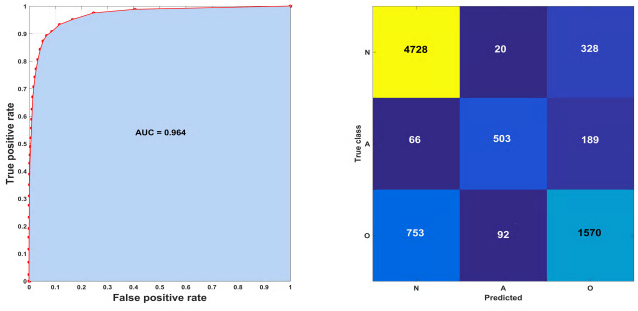

FIGURE 7. ROC curve and confusion matrix for best performance case.

The best performance is obtained for SNR = ∞ dB and all features involved. Figure 7 provided the cofusion matrix and ROC curve for this case.

IV. DISCUSSIONS

The performance metrics degrade slightly across almost all classification problems under different conditions (noise level

TABLE 3. Classification performances on whole dataset under different configurations for different tasks.

SNR	Features	Classfier	N/A				O/A				N/O				N/O/A			
			Acc	Sen	Spec	AUC	Acc	Sen	Spec	AUC	Acc	Sen	Spec	AUC	Acc	Sen	Spec	AUC
∞	All	SVM	0.957	0.756	0.987	0.976	0.896	0.697	0.959	0.924	0.834	0.626	0.933	0.874	0.804	0.656	0.982	0.957
		BT	0.966	0.823	0.988	0.987	0.895	0.697	0.957	0.939	0.849	0.677	0.931	0.891	0.821	0.652	0.982	0.964
	Top 10	SVM	0.963	0.827	0.983	0.972	0.896	0.749	0.943	0.922	0.803	0.608	0.896	0.840	0.782	0.730	0.973	0.957
		BT	0.962	0.829	0.981	0.977	0.895	0.751	0.941	0.942	0.810	0.630	0.895	0.855	0.786	0.721	0.971	0.960
	Top 8	SVM	0.963	0.828	0.983	0.972	0.896	0.748	0.943	0.921	0.804	0.606	0.898	0.837	0.782	0.731	0.972	0.956
		BT	0.963	0.835	0.982	0.978	0.894	0.755	0.938	0.938	0.806	0.621	0.894	0.849	0.783	0.724	0.970	0.960
	Top 5	SVM	0.959	0.816	0.980	0.969	0.885	0.697	0.944	0.914	0.796	0.593	0.892	0.833	0.772	0.691	0.972	0.952
		BT	0.959	0.822	0.979	0.975	0.881	0.734	0.927	0.930	0.793	0.605	0.883	0.833	0.766	0.703	0.965	0.955
30	All	SVM	0.958	0.758	0.988	0.974	0.896	0.700	0.957	0.918	0.830	0.651	0.915	0.872	0.803	0.662	0.982	0.955
		BT	0.965	0.818	0.987	0.985	0.893	0.694	0.956	0.935	0.846	0.672	0.929	0.890	0.818	0.650	0.982	0.962
	Top 10	SVM	0.963	0.827	0.983	0.971	0.895	0.731	0.946	0.921	0.799	0.600	0.894	0.838	0.779	0.718	0.974	0.955
		BT	0.961	0.826	0.981	0.978	0.895	0.740	0.943	0.939	0.807	0.622	0.895	0.853	0.784	0.716	0.972	0.961
	Top 8	SVM	0.963	0.827	0.983	0.971	0.895	0.734	0.946	0.920	0.800	0.597	0.897	0.836	0.778	0.717	0.973	0.954
		BT	0.961	0.830	0.981	0.980	0.893	0.746	0.939	0.936	0.805	0.617	0.895	0.847	0.782	0.718	0.970	0.960
	Top 5	SVM	0.961	0.825	0.982	0.972	0.886	0.695	0.946	0.914	0.794	0.582	0.895	0.832	0.769	0.690	0.972	0.955
		BT	0.960	0.830	0.979	0.977	0.879	0.720	0.929	0.928	0.791	0.596	0.883	0.830	0.763	0.693	0.963	0.956
20	All	SVM	0.957	0.751	0.988	0.973	0.895	0.691	0.959	0.918	0.830	0.657	0.912	0.871	0.802	0.651	0.983	0.954
		BT	0.963	0.805	0.987	0.983	0.894	0.693	0.957	0.932	0.846	0.671	0.929	0.888	0.817	0.647	0.982	0.959
	Top 10	SVM	0.960	0.816	0.982	0.969	0.889	0.727	0.940	0.913	0.797	0.599	0.892	0.836	0.774	0.708	0.971	0.952
		BT	0.962	0.826	0.982	0.976	0.893	0.738	0.942	0.936	0.806	0.618	0.895	0.851	0.782	0.707	0.972	0.958
	Top 8	SVM	0.960	0.816	0.981	0.968	0.884	0.721	0.936	0.912	0.797	0.595	0.893	0.833	0.773	0.705	0.970	0.951
		BT	0.962	0.831	0.981	0.976	0.888	0.740	0.934	0.932	0.802	0.608	0.894	0.844	0.776	0.713	0.968	0.956
	Top 5	SVM	0.958	0.809	0.980	0.970	0.882	0.695	0.941	0.907	0.794	0.580	0.896	0.830	0.769	0.679	0.969	0.950
		BT	0.960	0.831	0.979	0.974	0.877	0.725	0.925	0.922	0.791	0.600	0.881	0.833	0.762	0.699	0.962	0.949
10	All	SVM	0.948	0.656	0.991	0.959	0.884	0.619	0.967	0.901	0.813	0.562	0.933	0.849	0.784	0.558	0.986	0.935
		BT	0.950	0.692	0.989	0.969	0.872	0.604	0.957	0.908	0.827	0.618	0.927	0.868	0.791	0.546	0.984	0.940
	Top 10	SVM	0.942	0.666	0.983	0.956	0.867	0.594	0.952	0.887	0.788	0.532	0.910	0.825	0.756	0.540	0.979	0.935
		BT	0.944	0.703	0.981	0.963	0.692	0.651	0.938	0.907	0.794	0.577	0.897	0.834	0.762	0.607	0.972	0.937
	Top 8	SVM	0.942	0.661	0.984	0.952	0.855	0.527	0.958	0.867	0.785	0.514	0.914	0.820	0.749	0.510	0.978	0.926
		BT	0.942	0.694	0.979	0.960	0.853	0.605	0.930	0.896	0.784	0.563	0.889	0.828	0.750	0.567	0.969	0.929
	Top 5	SVM	0.924	0.575	0.977	0.937	0.824	0.511	0.922	0.842	0.781	0.509	0.911	0.813	0.734	0.433	0.969	0.909
		BT	0.942	0.639	0.967	0.947	0.815	0.566	0.893	0.868	0.773	0.549	0.880	0.811	0.725	0.482	0.951	0.909

and classifier configuration) when we eliminate up to 90% of the low-ranked features, which validates the effectiveness of the mutual information based, model-free feature selection scheme. Reduced computational cost and inference time can thus be anticipated in order to adapt to wearable scenarios. Generally speaking, no significant performance degradation was observed when we test our classifier on a subset which was not involved in the training procedure. The generalization ability of our method was validated.

For two-class classification problems, we have observed that very high accuracy 96.6% is obtained when only normal (N) and AF (A) recordings are involved, due to the sensitivity of nonlinear descriptors to arrhythmias. Under 10 dB configuration, the accuracies degrade slightly, maintaining a high level from 92.4% (using only top 5 features) to 95.0% (using all features). However, we have also observed that for other two class classification problems (N vs O and O vs A) the accuracy is off about 10%. Since we do not know the exact condition of patients, the internal high heterogeneity of other arrhythmia class (O) limited the performance especially the sensitivity in N vs O when O is taken as positive class.

For three-class classification problem, average accuracies up to 82.1% are obtained with high specificities. From the confusion matrix for the best case in 7 we found that a considerable part of the O class recordings are misclassified to be N class. It implies that some kinds of arrhythmias behave like normal rhythms with respect to the dynamics characterized by the descriptors we deploy. Since no more details within the O class are available, more targeted feature engineering is difficult.

Moreover, AFEv has reached a nearly perfect performance on benchmark dataset including MIT-BIH AF [27] without much room for further improvement. However, an improvement of its performance on new dataset like CinC 2017 is still possible (94.6% accuracy and 73% sensitivity are obtained under noiseless condition when using AFEv as the only predictor for N vs A task, and only 65% accuracy for N vs A vs O, not reported in Results). Since the benchmark datasets may be still 'small' and 'very clean', cannot cover the exact distribution of data encountered in real-world, hybrid feature may help to improve the robustness of detection scheme, but without the progress from medicine side or big enough data, the questioning about the generality of such machine learning algorithms will never cease.

The involvement of O class has raised interesting questions. For a wide population, we might not have enough prior knowledge about the distribution of certain diseases, neither the medical history of patients. AF and other arrhythmias are all likely to happen. Previous studies on AF detection reported the performance metrics obtained from datasets comprised of only normal recordings and AF recordings, usually the signal quality is also very good. The prior

TABLE 4. Classification performances on test set under different configurations for different tasks.

SNR	Features	Classfier	N/A			O/A			N/O			N/O/A		
			Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec
∞	All	SVM	0.953	0.743	0.985	0.891	0.675	0.959	0.827	0.625	0.923	0.797	0.632	0.982
		BT	0.966	0.832	0.986	0.890	0.693	0.953	0.847	0.681	0.925	0.818	0.646	0.981
	Top 10	SVM	0.962	0.826	0.982	0.897	0.746	0.944	0.803	0.616	0.892	0.782	0.730	0.973
		BT	0.959	0.837	0.977	0.896	0.762	0.938	0.807	0.635	0.889	0.782	0.720	0.967
	Top 8	SVM	0.961	0.823	0.982	0.897	0.747	0.945	0.803	0.610	0.894	0.780	0.732	0.972
		BT	0.961	0.840	0.979	0.893	0.761	0.935	0.806	0.628	0.891	0.780	0.732	0.967
	Top 5	SVM	0.957	0.813	0.979	0.886	0.698	0.945	0.791	0.591	0.886	0.766	0.691	0.971
		BT	0.959	0.839	0.977	0.877	0.737	0.922	0.788	0.607	0.874	0.767	0.713	0.961
30	All	SVM	0.958	0.763	0.988	0.900	0.726	0.955	0.827	0.649	0.912	0.802	0.686	0.980
		BT	0.966	0.832	0.986	0.893	0.705	0.953	0.847	0.683	0.925	0.820	0.693	0.979
	Top 10	SVM	0.964	0.836	0.983	0.896	0.750	0.942	0.800	0.597	0.897	0.780	0.735	0.972
		BT	0.963	0.833	0.983	0.898	0.763	0.940	0.806	0.622	0.894	0.785	0.728	0.969
	Top 8	SVM	0.964	0.835	0.983	0.896	0.750	0.942	0.799	0.593	0.898	0.779	0.733	0.971
		BT	0.963	0.837	0.982	0.893	0.761	0.935	0.805	0.614	0.896	0.782	0.718	0.969
	Top 5	SVM	0.963	0.837	0.982	0.888	0.715	0.942	0.790	0.577	0.892	0.768	0.708	0.970
		BT	0.963	0.846	0.981	0.878	0.741	0.921	0.789	0.594	0.881	0.762	0.707	0.961
20	All	SVM	0.953	0.743	0.984	0.895	0.697	0.957	0.819	0.645	0.902	0.793	0.657	0.981
		BT	0.963	0.822	0.984	0.897	0.704	0.959	0.838	0.663	0.922	0.811	0.660	0.981
	Top 10	SVM	0.959	0.824	0.979	0.886	0.725	0.936	0.787	0.589	0.881	0.764	0.710	0.969
		BT	0.962	0.840	0.980	0.901	0.749	0.949	0.798	0.606	0.889	0.775	0.727	0.970
	Top 8	SVM	0.959	0.825	0.979	0.885	0.730	0.934	0.787	0.585	0.883	0.763	0.708	0.968
		BT	0.961	0.841	0.979	0.891	0.749	0.935	0.795	0.595	0.890	0.772	0.735	0.967
	Top 5	SVM	0.958	0.823	0.977	0.880	0.697	0.937	0.785	0.560	0.892	0.758	0.681	0.967
		BT	0.960	0.841	0.977	0.876	0.718	0.926	0.784	0.590	0.876	0.756	0.704	0.961
10	All	SVM	0.948	0.661	0.991	0.876	0.601	0.963	0.812	0.562	0.931	0.779	0.545	0.985
		BT	0.950	0.697	0.988	0.871	0.610	0.953	0.827	0.615	0.928	0.790	0.530	0.985
	Top 10	SVM	0.938	0.649	0.982	0.859	0.582	0.946	0.785	0.528	0.907	0.751	0.512	0.977
		BT	0.944	0.707	0.979	0.876	0.659	0.945	0.797	0.588	0.896	0.764	0.611	0.971
	Top 8	SVM	0.937	0.647	0.981	0.850	0.526	0.952	0.782	0.511	0.911	0.746	0.506	0.975
		BT	0.941	0.697	0.977	0.847	0.611	0.922	0.780	0.559	0.886	0.750	0.549	0.967
	Top 5	SVM	0.920	0.568	0.972	0.816	0.509	0.913	0.777	0.504	0.907	0.729	0.428	0.963
		BT	0.920	0.627	0.964	0.809	0.577	0.882	0.771	0.551	0.876	0.719	0.495	0.945

distribution of benchmark datasets precludes some phenomenon which would always be inevitable if we do not go beyond and give up the pre-assumptions made by the benchmark datasets explicitly, like the existence and thus misclassification of arrhythmias other than AF. The real problem in application scenarios is to detect abnormal ECG recordings and classify them into one of a large number of arrhythmias, instead of *detecting AF from normal sinus rhythm*. Whatsmore, when it comes to the application in wearable scenarios, how can we decide to use the detection algorithms of one disease instead of another, or should we run all the algorithms customized for each specific disease and condition in parallel? At least we should take the internal differences between benchmark datasets and real-world data streaming into consideration.

However, we have also observed stable high specificity of the proposed algorithm. High specificity is a welcomed characteristic for preliminary screening of diseases because negative examples take an overwhelming portion in real world scenarios, so high specificity could be beneficial to lower the burden to medical system by excluding non-target population with a high degree of precision. And people who get alarms raised from wearables can get more accurate medical diagnosis from hospital. Generally speaking, we have found that even for relatively short and noise-contaminated ECG recordings, nonlinear descriptors of HRV are still seem to be robust and efficient. Since R peak is the most dominant fiducial point in the morphology of ECG and the last characteristic wave to be submerged by severe interference, the lower signal to noise ratio (SNR) for the recording as a whole may not necessarily lead to a worse description about the heart activities extracted from RRI. Heart rate variability based features are thus more stable.

V. CONCLUSION

An automatic atrial fibrillation detection algorithm was proposed aiming for detecting AF from short ECG recordings since high quality long term recording is not always guaranteed in wearable scenarios for diagnosis where the measurement is always exposed to different kinds of noises and artifacts. The algorithm was validated on a recently released dataset comprised of short ECG recordings acquired by non-standard portable device in non-structured environments. Good accuracy and very high specificity of our algorithm made it a competitive algorithm prototype for preliminary screening. Our research supported the effectiveness of heart rate variability based nonlinear features in AF detection from short ECG recording. Considering the existence of arrhythmias other than AF, cascading classification can be a direction of future work.

REFERENCES

[1] Y.-K. Iwasaki, K. Nishida, T. Kato, and S. Nattel, "Atrial fibrillation pathophysiology: Implications for management," *Circulation*, vol. 124, no. 20, pp. 2264–2274, Nov. 2011.

[2] N. S. Peters, R. J. Schilling, P. Kanagaratnam, and V. Markides, "Atrial fibrillation: Strategies to control, combat, and cure," *Lancet*, vol. 359, no. 9306, pp. 593–603, Feb. 2002.
[3] S. K. Sahoo, W. Lu, S. D. Teddy, D. Kim, and M. Feng, "Detection of Atrial fibrillation from non-episodic ECG data: A review of methods," in *Proc. EMBC*, Aug./Sep. 2011, pp. 4992–4995.
[4] E. Ros, S. Mota, F. J. Fernández, F. J. Toro, and J. L. Bernier, "ECG characterization of paroxysmal atrial fibrillation: Parameter extraction and automatic diagnosis algorithm," *Comput. Biol. Med.*, vol. 34, pp. 679–696, Dec. 2004.
[5] M. Malik *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, Mar. 1996.
[6] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang, "Automatic Online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy," *BioMed. Eng. OnLine*, vol. 13, no. 1, p. 18, 2014.
[7] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals," *Med. Biol. Eng. Comput.*, vol. 39, no. 6, pp. 664–671, Nov. 2001.
[8] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in Poincare plot," *Biomed. Eng. OnLine*, vol. 8, no. 1, p. 38, 2009.
[9] S. Bouwstra, W. Chen, L. Feijs, and S. B. Oetomo, "Smart jacket design for neonatal monitoring with wearable sensors," in *Proc. 6th Int. Workshop Wearable Implant. Body Sensor Netw. (BSN)*, Berkeley, CA, USA, Jun. 2009, pp. 162–167.
[10] C. Zou, Y. Qin, C. Sun, W. Li, and W. Chen, "Motion artifact removal based on periodical property for ECG monitoring with wearable systems," *Pervasive Mobile Comput.*, vol. 40, pp. 267–278, Sep. 2017.
[11] H. Chen, Y. Zhao, Z. Mei, Y. Mei, S. B. Oetomo, and W. Chen, "Characterization of a novel carbonized foam electrode for wearable bio-potential recording," in *Proc. IEEE 15th Int. Conf. Wearable Implant. Body Sensor Netw.*, Las Vegas, NV, USA, Mar. 2018, pp. 173–176.
[12] M. A. Yokus and J. S. Jur, "Fabric-based wearable dry electrodes for body surface biopotential recording," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 2, pp. 423–430, Feb. 2016.
[13] B. Yang, C. Yu, and Y. Dong, "Capacitively coupled electrocardiogram measuring system and noise reduction by singular spectrum analysis," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3802–3810, May 2016.
[14] W. Chen, J. Hu, S. Bouwstra, S. B. Oetomo, and L. Feijs, "Sensor integration for perinatology research," *Int. J. Sensor Netw.*, vol. 9, no. 1, pp. 38–49, 2011.
[15] B. Taji, S. Shirmohammadi, V. Groza, and I. Batkin, "Impact of skin-to-electrode interface on electrocardiogram measurements using conductive textile electrodes," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1412–1422, Jun. 2014.
[16] D. E. Albert and B. R. Satchwell, "Universal ECG electrode module for smartphone," U.S. Patent 20 150 073 285 A1, Mar. 12, 2015.
[17] J. Smith, J. Ward, T. Urbano, and M. Mueller, "Use of alivecor heart monitor for heart rate and rhythm evaluation in dairy water buffalo calves," *J. Dairy, Veterinary Animal Res.*, vol. 4, no. 2, p. 00113, Nov. 2016.
[18] *The PhysioNet/Computing in Cardiology Challenge*, Nat. Inst. Gen. Med. Sci. (NIGMS), Nat. Inst. Biomed. Imag. Bioeng. (NIBIB), 2017, [Online]. Available: https://physionet.org/challenge/2017/
[19] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
[20] G. E. Billman, "Heart rate variability—A historical perspective," *Frontiers Physiol.*, vol. 3, p. 86, Nov. 2011.
[21] G. B. Moody, "Spectral analysis of heart rate without resampling," in *Proc. Comput. Cardiol. Conf.*, Sep. 1993, pp. 715–718.
[22] J. D. Scargle, "Studies in astronomical time series analysis. II—Statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, Dec. 1982.
[23] E. M. Stein and R. Shakarchi, *Fourier Analysis: An introduction*. Princeton, NJ, USA: Princeton Univ. Press, 2010.
[24] Z. Mei, X. Zhao, H. Chen, and W. Chen, "A distributed descriptor characterizing structural irregularities of EEG time series for epileptic seizure detection," in *Proc. EMBC*, 2018.
[25] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?" *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1342–1347, Nov. 2001.
[26] C. K. Karmakar, A. H. Khandoker, J. Gubbi, and M. Palaniswami, "Complex correlation measure: A novel descriptor for Poincaré plot," *BioMed. Eng. OnLine*, vol. 8, no. 1, p. 17, 2009.
[27] S. Sarkar, D. Ritscher, and R. Mehra, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1219–1224, Mar. 2008.
[28] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, no. 2, pp. 70–73, Jun. 1967.
[29] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[30] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.
[31] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handbook Statist.*, vol. 24, pp. 303–329, 2005, doi: 10.1016/S0169-7161(04)24011-1.
[32] P. Rajpurkar, A. Y. Hannun, M. Haghpanah, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," in *Proc. Comput. Vis. Pattern Recognit.*, 2017. [Online]. Available: http://arxiv.org/abs/1707.01836

ZHENNING MEI received the B.S. degree in electronics engineering from Fudan University, Shanghai, China, in 2012, where he is currently pursuing the M.S. degree in circuits and systems. His research interests include signal processing and pattern recognition.

XIAO GU received the B.E. degree in electronics engineering from Fudan University, Shanghai, China, in 2018. His research interests include computer vision and pattern recognition.

HONGYU CHEN received the B.A. degree from Nanchang University and the M.S. degree in computer science from Zhejiang University. He is currently pursuing the Ph.D. degree with the Department of Industrial Design, Technical University of Eindhoven. His research interests include neonatal cardiovascular monitoring.

WEI CHEN received the B.S. and M.S. degrees from Xian Jiaotong University in 1999 and 2002, respectively, and the Ph.D. degree from The University of Melbourne in 2007. From 2008 to 2015, she was an Assistant Professor with the Department of Industrial Design at the Technical University of Eindhoven. She is currently a Full Professor with the Department of Electronic Engineering, Fudan University. Her research interests include wearable sensor system, biomedical signal processing, and health monitoring. She is also an Associate Editor of the IEEE JOURNAL OF BIOMEDICAL HEALTH INFORMATICS, a Regional Representative of the IEEE EMBS Technical Committee on Wearable Biomedical Sensors and Systems, and the Vice Chair of the IEEE Sensors and Systems Council.