

PAPER

Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost

To cite this article: Yao Chen *et al* 2018 *Physiol. Meas.* **39** 104006

View the [article online](#) for updates and enhancements.

You may also like

- [Hi-gMISnet: generalized medical image segmentation using DWT based multilayer fusion and dual mode attention into high resolution pGAN](#)
Tushar Talukder Showrav and Md Kamrul Hasan
- [An open source benchmarked toolbox for cardiovascular waveform and interval analysis](#)
Adriana N Vest, Giulia Da Poian, Qiao Li et al.
- [Probabilistic model-based approach for heart beat detection](#)
Hugh Chen, Yusuf Erol, Eric Shen et al.



PAPER

Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost

RECEIVED
14 February 2018REVISED
7 August 2018ACCEPTED FOR PUBLICATION
5 September 2018PUBLISHED
24 October 2018Yao Chen^{1,2}, Xiao Wang^{1,2}, Yonghan Jung^{1,3}, Vida Abedi⁴ , Ramin Zand⁴, Marvi Bikak⁵ and Mohammad Adibuzzaman¹¹ Regenstrief Center for Healthcare Engineering, Purdue University, West Lafayette, IN, United States of America² Department of Statistics, Purdue University, West Lafayette, IN, United States of America³ Department of Industrial Engineering, Purdue University, West Lafayette, IN, United States of America⁴ Geisinger Medical Center, Danville, PA, United States of America⁵ Indiana University School of Medicine, Indianapolis, IN, United States of AmericaE-mail: chen2037@purdue.edu, wangxiao@purdue.edu, jung222@purdue.edu, vabedi@geisinger.edu, rzand@geisinger.edu, mbikak@iu.edu and madibuzz@purdue.edu**Keywords:** atrial fibrillation, electrocardiography, piecewise linear function, XGBoost

Abstract

Objective: Detection of atrial fibrillation is important for risk stratification of stroke. We developed a novel methodology to classify electrocardiograms (ECGs) to normal, atrial fibrillation and other cardiac dysrhythmias as defined by the PhysioNet Challenge 2017. **Approach:** More specifically, we used piecewise linear splines for the feature selection and a gradient boosting algorithm for the classifier. In the algorithm, the ECG waveform is fitted by a piecewise linear spline, and morphological features relating to the piecewise linear spline coefficients are extracted. XGBoost is used to classify the morphological coefficients and heart rate variability features. **Main results:** The performance of the algorithm was evaluated by the PhysioNet Challenge database (3658 ECGs classified by experts). Our algorithm achieved an average F_1 score of 81% for a 10-fold cross-validation and also achieved 81% for F_1 score on the independent testing set. This score is similar to the top 9th score (81%) in the official phase of the PhysioNet Challenge 2017. **Significance:** Our algorithm presents a good performance on multi-label short ECG classification with selected morphological features.

1. Introduction

Atrial fibrillation (AF) is an abnormal heart rhythm characterized by rapid and irregular heartbeat. AF, usually associated with significant mortality and morbidity, is the most common sustained cardiac arrhythmia, occurring in 1%–2% of the general population (Camm *et al* 2010, Lip *et al* 2016). At least 2.7 million Americans are living with AF, and more than 12 million Europeans and North Americans are estimated to suffer from AF (Camm *et al* 2010, Colloca *et al* 2013). The incidence of AF increases with age, from less than 0.5% at 40–50 years, to 5%–15% for 80 years of age (Naccarelli *et al* 2009). Its prevalence will likely triple in the next 30–50 years, particularly in the United States and other western countries with aging population demographics (Savelieva and Camm 2008). This growth may also be influenced by extended survival outcomes for patients with congestive heart failure, valvular heart disease, and coronary artery disease, as AF is common among patients with other forms of structural heart disease.

Accurate diagnosis of AF is the first step to address this problem and is essential in mitigating such serious concerns. There have been many previous research studies related to the classification of abnormal ECG beats for arrhythmia (Übeyli 2007, Übeyli 2009, Daqrouq *et al* 2014, Jung and Kim 2017, Afdala *et al* 2017, Rajpurkar *et al* 2017). Most of the proposed models have used classical classifiers such as support vector machines (SVMs) or neural network with discrete wavelet transformation of the signals to capture the morphological characteristics of the ECG (Daqrouq *et al* 2014, Rajpurkar *et al* 2017). One of the key challenges for AF detection is that it can be

episodic, hence AF classification from continuous monitoring data in real settings is important for its accurate diagnosis. With the advent of wearable monitoring devices and increased computational power, researchers have endeavored to develop comparable classification methods in more realistic settings. For example, Sankari and Adeli (2011) tackled the classification of ECG beats collected from mobile devices to identify AF and myocardial infarction. Due to the limited computational capacity of those mobile devices, the classification model was developed based on heart rate variability, instead of analyzing the morphological features of ECG beats, therefore under-utilizing rich temporal data. In another study, the classification problem in a wearable device environment was studied; however, only the compressed-sensed ECG signals (encoded by discrete wavelet transformation and basis pursuit denoising) were considered (Da Poian *et al* 2017). With similar constraints on computational power, the SVM method was employed on heart rate variability. To make the algorithm computationally efficient, Mohebbi and Ghassemian (2011) proposed a method focusing on heart rate variability without dealing with high dimensional morphological features.

We tackled the AF classification problem based on ECG recordings (PhysioNet Challenge 2017 data (PhysioNet 2017)) collected from an AliveCor device, which is an ECG recording device in the mobile environment (Clifford *et al* 2017). The proposed model can capture heart rate variability and morphological features without generating high dimensional features as wavelet analysis does. To circumvent representing the morphology in high dimension, we employed a signal fitting method called *piecewise linear function*. XGBoost (Chen and Guestrin 2016), a gradient-boosting-method-based classifier known for its performance in many data analysis competitions, was used to improve classification performance. We achieved an F_1 score of 81% on the test set from the PhysioNet Challenge, which is comparable to other high-ranked competitive classifiers (Clifford *et al* 2017).

Overall, our approach focuses on identification of features from waveform morphology with piecewise linear splines and (i) generates fewer number of features than the methods based on the discrete wavelet transformation, making this method more efficient and statistically robust; (ii) can be applied to any type of ECG recordings; in comparison, most existing methods are trained on standard ECG databases collected in hospitals and not contaminated by any external noises; they also focus only on classification between normal and AF rhythms; and (iii) uses the Kaggle Inc.⁶ winning algorithm XGBoost for the classifier; this approach is highly efficient and flexible and can be easily used on distributed platforms for further computational efficiency.

2. Methods

2.1. Challenge data

ECG recordings, collected and band-pass filtered using an AliveCor device, were sampled at 300 Hz. The training set contains 8528 single-lead ECG recordings ranging from 9 s to just over 60 s. The test set (withheld by the organizers) contains 3658 ECG recordings of similar lengths. Each recording is labeled as either ‘Normal’, ‘AF’, ‘Other’ or ‘Noisy’. All labelling was performed by a single expert. During the various phases of the challenge, reference labels for the training data of the ECGs were updated with three reference versions provided by the organizers. The data profile in table 1 is given based on the latest version. The test set was unavailable to the public and was not accessible to us before the submission of this manuscript.

2.2. AF classification algorithm overview

AF is defined as a ‘tachyarrhythmia characterized by predominantly uncoordinated atrial activation with consequent deterioration of atrial mechanical function’ by the American College of Cardiology (ACC), the American Heart Association (AHA) and the European Society of Cardiology (ESC) (Fuster *et al* 2001).

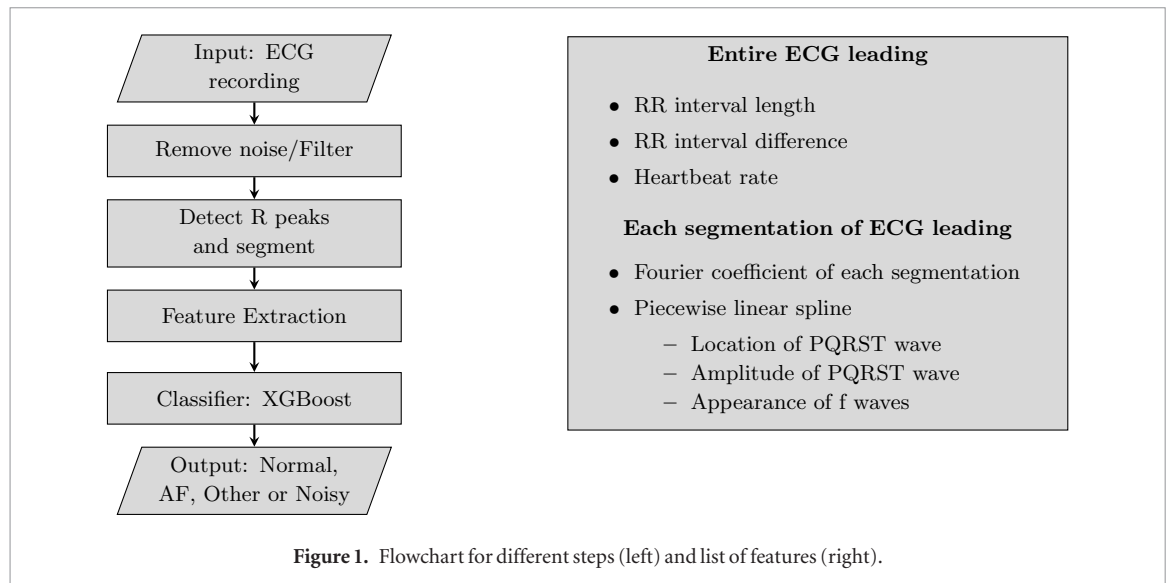
Despite the enormity of this issue, AF detection remains problematic, as it can be episodic. The irregular rhythms in the ECG, which can be captured by the underlying pattern of R waves, is a key factor when diagnosing AF. Another important factor is the absence of P waves. AF detection by an algorithm can be considered as one of two schools of thought: an atrial-activity-analysis-based or ventricular-response-analysis-based method. Atrial-activity-analysis-based AF detectors are based on the analysis of the absence of P waves or the presence of fibrillatory f waves in the TQ interval. In contrast, ventricular response analysis is based on the predictability of the inter-beat timing (‘RR interval’) of the QRS complexes in the ECG (PhysioNet 2017) (figure 2). We have developed a hybrid method, where both approaches are combined for the selection of features. To extract information according to atrial activity and ventricular response, we break the method into several steps, which are shown in the flowchart (figure 1).

We implemented the algorithm in Python 2.7, and incorporated some existing packages, ‘biosppy’ (Carreiras 2015) and ‘scipy’ (Jones *et al* 2001) for pre-processing, such as denoising and re-sampling for the wavelet method. The model was trained on the training dataset and stored as a separate file. To predict a new record, the

⁶ A machine learning challenge: www.kaggle.com/.

Table 1. Data profile for the training set.

Type	# Recording	Time length (s)				
		Mean	SD	Max	Median	Min
Normal	5076	32.11	9.97	60.95	30.0	9.05
AF	758	32.34	12.32	60.21	30.0	9.99
Other	2415	34.30	11.76	60.86	30.0	9.13
Noisy	279	24.38	10.41	60.0	30.0	9.36
Overall	8528	32.50	10.89	60.95	30.0	9.05

**Figure 1.** Flowchart for different steps (left) and list of features (right).

model takes an ECG recording as an input and returns a class label of either ‘Normal’, ‘AF’, ‘Other’ or ‘Noisy’. The evaluation was performed by running the algorithm on the server, equipped with virtual machines (VMs), provided by the challenge organizers. Each VM is configured with a single-core AMD 64 processor, 2 GB of RAM, a 2 GB read-write/home partition, and a 500 MB read-write/tmp partition. Each classification task was limited to 2×10^{11} CPU instructions⁷. Our final submission consisted of 34 recordings that exceeded the computational limit and were classified as ‘Noisy’.

2.3. R peaks detection and PQRST segmentation

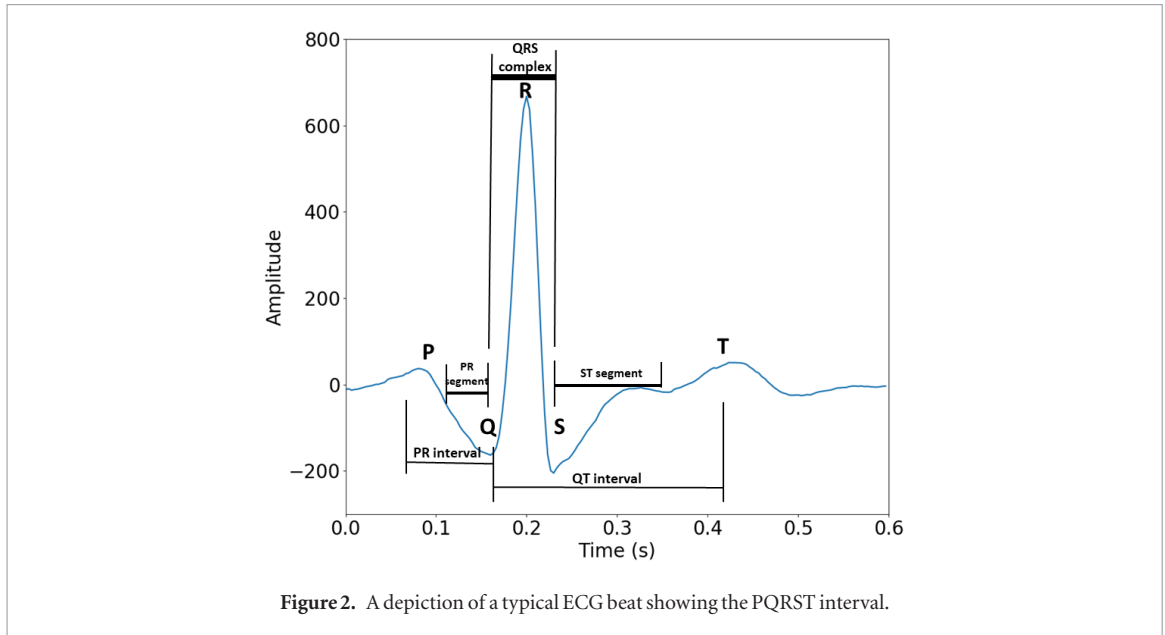
A complete normal heartbeat produces four entities on ECG: a P wave, a QRS complex, a T wave and a U wave, where the U wave is not typically seen and its absence is generally ignored. Therefore, a PQRST interval, as shown in figure 2, is considered to represent a complete heartbeat wave on an ECG recording. Depending on the source lead of the ECG, this interval might be inverted (i.e. negative R peaks) for some waves.

R peaks detection and PQRST segmentation is the first step for analyzing the ECG and classifying different categories. Correctly identifying the R peaks and making proper PQRST segmentation could provide valuable information regarding different features including heart rate and RR interval, which can also facilitate the discovery of the irregularity of the heart rhythm.

2.3.1. Comparison of different methods for R peaks detection

Different methods have been proposed to identify R peaks. We compared five of the existing methods including Christov (2004), Engelse and Zeelenberg (1979), slope sum function (SSF) Zong *et al* (2003), Hamilton (2002) and Gamboa (2008). These five methods have shown an accuracy of 90% or more for QRS detection on the PhysioNet database (Goldberger *et al* 2000, Canento *et al* 2013). For the challenge dataset, the performance of these five methods does not differ significantly from each other in the case of positive R peaks (Canento *et al* 2013). However, Hamilton (2002) method performed better for automatically detecting the negative R peaks.

⁷www.physionet.org/challenge/sandbox/.



2.4. Fitting heartbeat

Both atrial-activity-analysis-based or ventricular-response-analysis-based methods require capturing the morphological features of the ECG wave. Traditional methods use wavelet transformation, and then consider the coefficients as a representation of the ECG for further analysis. However, it is difficult to extract the P, Q, R, S and T locations and amplitude information from the wavelet coefficients, since the PQRST signals vary their positions randomly (Nair *et al* 2013). We introduce piecewise linear splines for capturing the waveform morphology at the heartbeat level.

2.4.1. Piecewise linear function

Piecewise linear functions, such as adaptive piecewise estimation, are commonly used in non-parametric studies to fit a function (Tibshirani *et al* 2014). The method uses a series of end-to-end straight lines to approximate the wave or function. The location of end points on the X axis are called knots. The goal is to minimize the sum of the least squared errors between the piecewise linear function and the true function to achieve a better fit. Figure 3 shows a simple example of a piecewise linear spline (red) for estimating a quadratic function whose parameters are unknown (blue). The mathematical form of the piecewise linear spline is shown in equation (1). We define $f(x)$ to be the function on $[0, 1]$ without loss of generality:

$$f(x, \mathcal{B}) = \beta_0 + \beta_1 x + \sum_{i=1}^k \beta_{i+1} (x - t_i)_+ \quad (1)$$

where $\mathbf{t} = \{0 = t_0 < t_1 < \dots < t_{k+1} = 1\}$ are the knots we chose, and $\mathcal{B}$ s are the coefficients for each spline.

The keys to a good approximation on a piecewise linear model are (i) the number of knots, and (ii) the location of the knots. With the number of knots and the location of the knots fixed, our question can be simplified to an optimization problem with a quadratic loss function, as in equation (2):

$$L(y, f(x, \mathcal{B})) = \sum_{i=1}^N (y - f(x, \mathcal{B}))^2 \quad (2)$$

where $f(x, \mathcal{B})$ is defined in equation (1). Our task is simplified to minimize equation (2) on the space of $\mathcal{B}$, in equation (3):

$$\hat{\mathcal{B}} = \operatorname{argmin}_{\mathcal{B}} L(y, f(x, \mathcal{B})). \quad (3)$$

Here, we propose a forward step-wise algorithm and adaptively add a new knot to the current knots in the most likely position iteratively (Wang and Liu 2018). The details of each step are given in algorithm 1. Two examples of the fitting of piecewise linear spline are shown in figure 4. The examples include fitting the function for a Doppler function and a PQRST segment of a heartbeat.

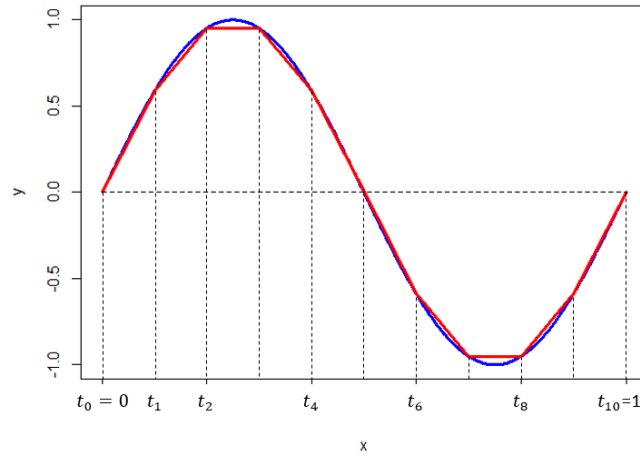


Figure 3. Piecewise linear spline example. Here a function $f(x)$ (blue) is estimated by a piecewise linear spline (red) when the knots $\{t_0 = 0, \dots, t_{10} = 1\}$ are given.

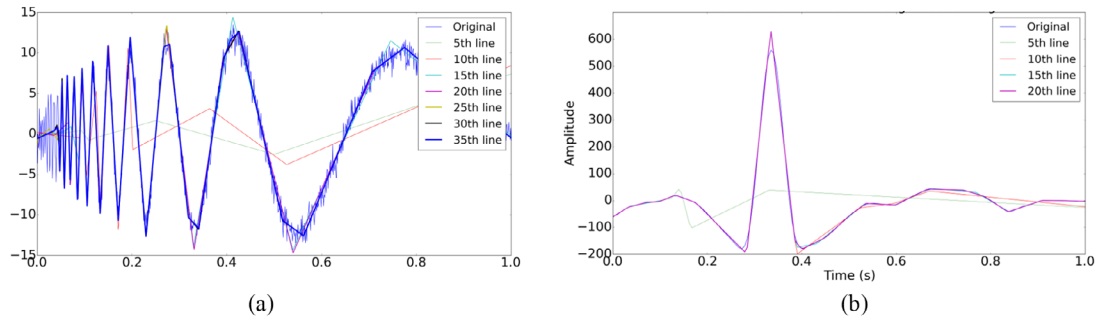


Figure 4. The examples show the approximation of functions and waves with piecewise linear functions. For each heartbeat, it takes at most 30 iterations to find a well-fitted piecewise linear function. (a) Fitting the Doppler function with piecewise linear function. (b) Fitting a heartbeat with piecewise linear function.

Algorithm 1. Fitting piecewise linear functions (Wang and Liu 2018).

```

1 Given significant level  $\alpha$ ,  $f(x_i)$   $i = 1, \dots, n$ ;
2 for  $k = 1, \dots, \text{do}$ 
3   Find  $\{t_1, \dots, t_k\} \in [t_0 = 0, t_{k+1} = 1]$  and corresponding  $\hat{f}$  so that it minimize equation (2);
4   Set residuals  $\mathcal{R}_l = \{f(x_j) - \hat{f}(x_j) | x_j \in [t_{l-1}, t_l]\}$ , for  $l = 1, \dots, k$ ;
5   Test the independence on  $\mathcal{R}_l$  at significant level  $\alpha$ ,  $l = 1, \dots, k$ , collect p-value  $p_l$  for each set  $\mathcal{R}_l$ ;
6   if  $\min\{p_i\} < \alpha$  then
7     continue;
8   else
9     Residuals are independent in all sets at significant level of  $\alpha$ ;
10    break;
11  end
12 end

```

The piecewise linear function approximated from the algorithm helps us to correctly identify the position of the peaks in the ECG, which is critical for the analysis of the atrial activity in detecting the absence of P waves and the presence of f waves by capturing the morphology.

2.5. Feature extraction

To detect AF, clinicians seek the rhythm of the heartbeat, absence of P waves and presence of f waves (figure 5). An F wave is an atrial flutter wave on the ECG, which is more of a regular tachyarrhythmia and often superimposed with atrial fibrillation. It appears as a ‘sawtooth’ pattern in leads II, III and aVF (figure 5).

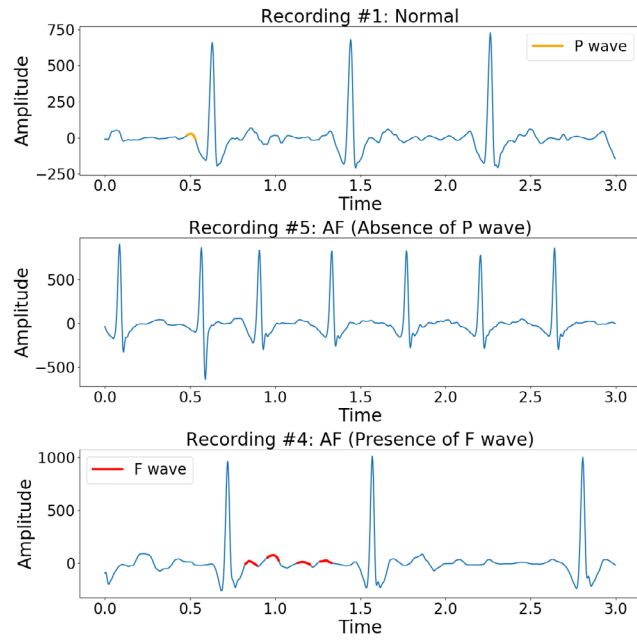


Figure 5. Examples of ECGs of atrial fibrillation. Top: The arrow indicates a P wave in normal sinus rhythm, which is lost in atrial fibrillation (middle); bottom: the morphology of f waves in ECG. The irregularity of heartbeat can be seen in both graphs.

We have developed a set of rules to extract the statistical features of heartbeat rate for the ECG signal and to detect the P wave- and f wave-related features from the coefficients of the heartbeat approximation based on a piecewise linear spline. These rules are described in detail in this section.

2.5.1. Heartbeat rate and rhythm

The detection of R peaks and PQRST segmentation help us count the number of heartbeat over a period of time. The heartbeat rate of each ECG is tracked along with number of R peaks in a fixed time window. With the window shifting over the entire ECG, we get a sequence of heartbeats which provides the heart rate and the heart rate variability (HRV). We define the set $\{R_i \mid i = 1, \dots, n\}$ ⁸ to be the sequence of R peak locations of the ECG leading record, and heartbeat rate is defined as the number of R peaks detected over one second. We compute the length of RR intervals, which are the time differences between two adjacent R peaks, as $\{RR_i = R_{i+1} - R_i \mid i = 1, \dots, n-1\}$ and calculate the difference between RR interval lengths as $\{diff_i = RR_{i+1} - RR_i \mid i = 1, \dots, n-2\}$. Statistical measurements of $\{RR_i\}$, $\{diff_i\}$ such as mean and standard deviation are extracted, and the variation of the sequence indicative of the rhythm of the heartbeat.

2.5.2. Absence of P waves

The absence of P waves is captured by first annotating each wave in the PQRST intervals. After fitting the PQRST intervals to piecewise linear functions, we mark the inflection points of the piecewise linear functions as our candidates for each peak. Since R peaks are located by the first step, according to the characteristic of each wave, we annotate the PQRST peaks among the candidates according to their amplitude and distance to the R peaks, as shown in figure 6. With the annotation of PQRST, we can detect the absence of a P wave by identifying the presence of an inflection point before the Q wave.

2.5.3. Presence of f waves

The presence of f waves in the ECG is another measure used by clinicians for detection of AF. It usually appears as a ‘sawtooth’ pattern (figure 5). To detect AF, we at first identify the coefficients of the piecewise linear function and then find the inflection points. This step is similar to the detection of the absence of P waves. After that, we compute the number of inflection points that follows the R peak. If there is an irregular number of inflection points after the R wave, it is more likely that there is a ‘sawtooth’ pattern for the presence of f waves. We compute the proportion of such waves among all PQRST intervals in a single ECG leading and, depending on the percentage, we detect the f waves.

2.5.4. Other features

We also used other general features such as the RR interval, and the differences in RR interval and heartbeat rate, as described in figure 1. The argument for using differences in RR interval can be explained by a Lorenz plot

⁸ n denotes the number of R peaks detected in each ECG recording, and will vary for different recordings.

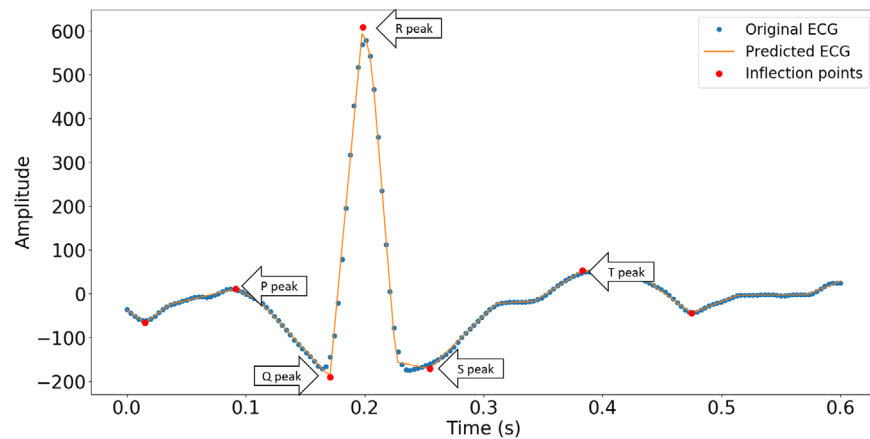


Figure 6. Annotation of PQRST wave from fitted piecewise linear function.

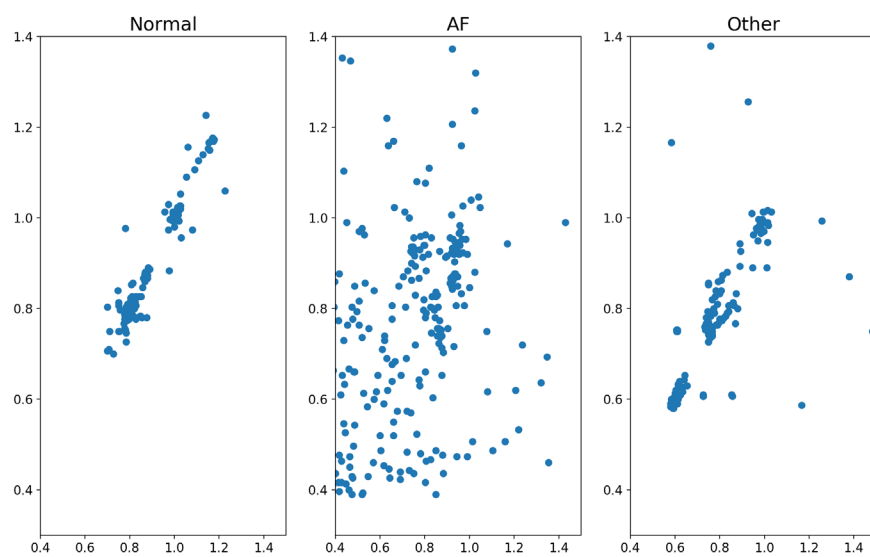


Figure 7. Lorentz plot showing the comparison between the differences of RR intervals for Normal, AF and Other classes for three samples. X axes are $RR(n) - RR(n-1)$, and Y axes are $RR(n+1) - RR(n)$ for each n , where n is an index of RR intervals.

(Hnatkova *et al* 1995) (figure 7). The slope gets closer to 1 if the RR intervals are regular without much variation; in contrast, it becomes more random if the RR intervals change irregularly, as shown in figure 7.

2.6. Classification

The features extracted using the piecewise linear function as well as other features that had been reported in the literature to classify AF, as described in the previous sections, were used for inputs for a classifier. The list of the features are (i) heart beat rate and rhythm, (ii) presence or absence of *P* waves as identified by piecewise linear function, (iii) presence or absence of *f* waves, (iv) RR interval and (v) differences in RR interval. For each sample, these features are extracted and a high-dimensional vector was created. These vectors were then split randomly for training and testing with a 10-fold cross-validation. We used 82% (7000) of the challenge dataset as our training data and the remaining 18% (1528) as our validation set. XGBoost, a gradient boosting algorithm, was used as the classifier of the ECG signal from the extracted features (Chen and Guestrin 2016). Other methods such as neural network (NN) were also explored. XGBoost, short for ‘extreme gradient boosting’, uses gradient-boosted trees to solve the problem of supervised learning. Gradient-boosted trees use decision trees of a fixed maximum size as base learners, and iteratively learns the base learners to add up to a final strong classifier (Chen and Guestrin 2016). A Python open source package for XGBoost was used for the implementation⁹.

2.7. Evaluation metrics

The results are measured in terms of F_1 score and receiver operating characteristics (ROC) curve, described in this section. The definition of F_1 score is based on a confusion matrix for the reference class and the predicted class. The confusion matrix is shown in table 2.

⁹<https://github.com/dmlc/xgboost>.

Table 2. Counting rules for the numbers of the variables.

		Predicted classification				
		Normal	AF	Other	Noisy	Total
Reference classification	Normal	Nn	Na	No	Np	$\sum N$
	AF	An	Aa	Ao	Ap	$\sum A$
	Other	On	Oa	Oo	Op	$\sum O$
	Noisy	Pn	Pa	Po	Pp	$\sum P$
Total		$\sum n$	$\sum a$	$\sum o$	$\sum p$	

The performance of the algorithm was evaluated as an F_1 measure as defined by the challenge organizers, which is an average of the three F_1 values from each classification type. F_1 values are defined with the following equations for the three categories.

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

The final score is calculated as

$$F_1 = (F_{1n} + F_{1a} + F_{1o})/3.$$

We also utilised other traditional statistical measures, such as ROC curves, precision-recall curves (PRC), positive predictive value (PPV) as well as sensitivity (a.k.a true positive rate) and specificity. The definitions of these parameters are listed below.

$$\begin{aligned} \text{Specificity} &= \frac{\text{Number of true negative}}{\text{Number of actually negative samples}} \\ \text{Sensitivity} &= \frac{\text{Number of true positive}}{\text{Number of actually positive samples}} \\ \text{PPV} &= \frac{\text{Number of true positive}}{\text{Number of positive calls}} \\ \text{False positive rate} &= \frac{\text{Number of false positive}}{\text{Number of actually negative samples}}. \end{aligned}$$

The ROC curve and PRC are graphical plots that illustrate the diagnostic performance of a binary classifier as its discrimination threshold is varied. A ROC curve is created by plotting the true positive rate against the false positive rate at various thresholding settings. A PRC is created by plotting the PPV against sensitivity, and shows the trade-off between precision and recall for different thresholds.

3. Results

3.1. Classification performance on the test set (hidden)

To build our model on the training data, we performed 10-fold cross-validation by randomly selecting 7000 (82%) records as the training set and the remaining 1528 (18%) records as the validation set. For the 10-fold cross-validation, we achieved an average F_1 score of 80.5%, and an accuracy of 83.8% with specificity of 98.3%. Among the F_1 scores, the Normal class, AF class and Other class had a score of 0.90, 0.78 and 0.74, respectively. Notably for the ‘AF’ class, we achieved a sensitivity of 0.78 and positive predictive value of 0.81. For the test set that was hidden from us, we achieved similar performance statistics. The detailed F_1 scores for the hidden test set are shown in table 3.

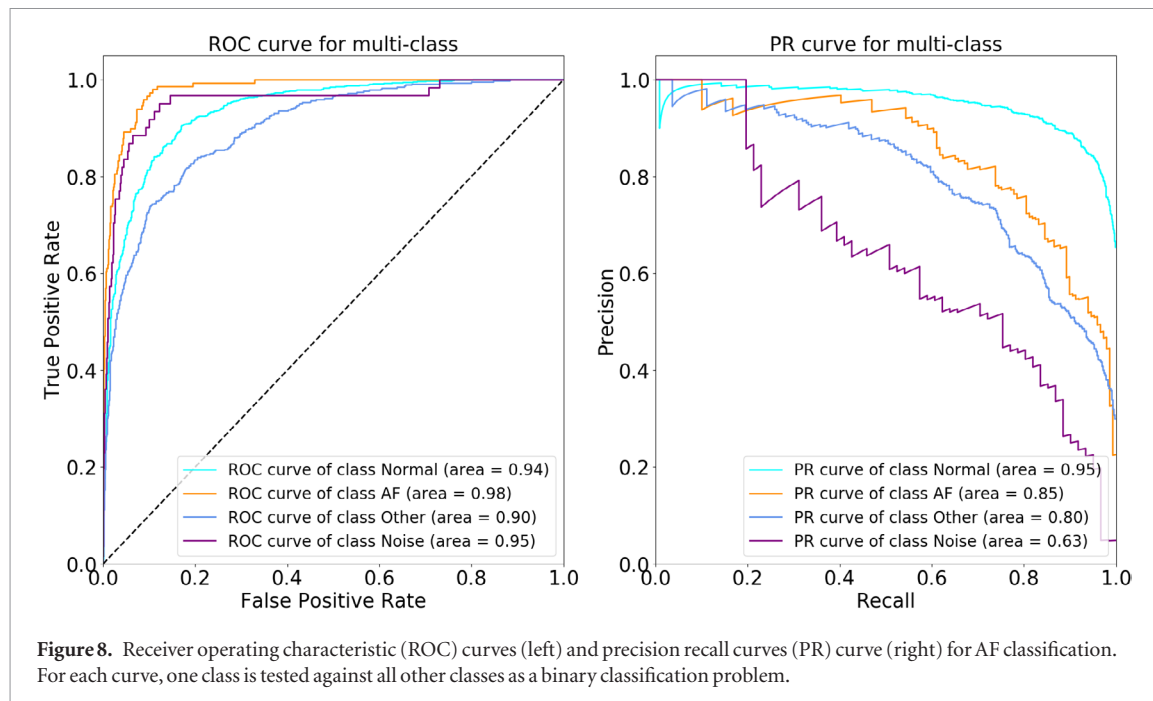
3.2. Precision recall (PR) curve and receiver operating characteristic (ROC) curve

For a multi-class classification problem, we characterize a ROC curve for each class, by testing that class against all other classes. We used 82% (7000) of the challenge dataset as our training data and the remaining 18% (1528) as our validation set to generate the ROC curves, shown in figure 8.

The ROC curves indicate good performance of the algorithm, when we consider the classification problem for each class as a binary classifier. We also studied the head-to-head comparison between each pair of classes and achieved the lowest area under the curve (AUC) for the Normal-Other pair, which is still 0.94. The AUC for each

Table 3. Final F_1 scores of piecewise linear spline/XGBoost method for the test set (hidden).

Class	F_1 score (%)
Normal	0.90
AF	0.80
Other	0.72
Overall	0.81

**Figure 8.** Receiver operating characteristic (ROC) curves (left) and precision recall curves (PR) curve (right) for AF classification. For each curve, one class is tested against all other classes as a binary classification problem.

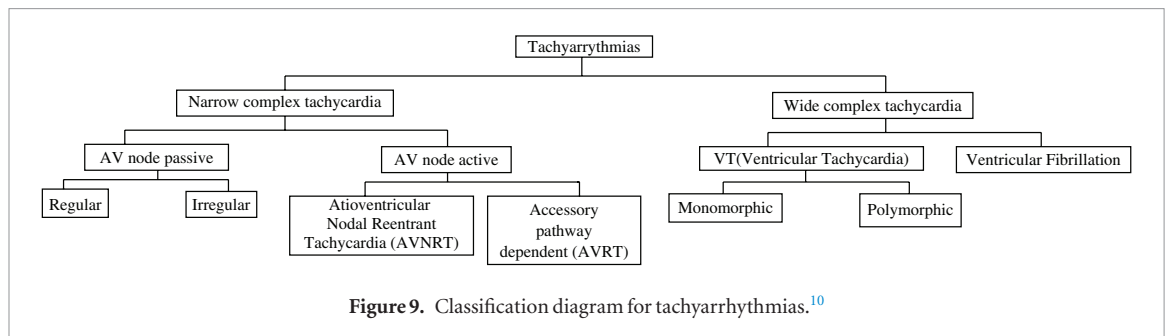
class is above 0.9, and the class 'AF' achieved an AUC of 0.98. As the data are imbalanced and a precision recall curve (PR) better represents the performance in such cases, we used a PR curve. The AUCs of the three major classes (Normal, AF, Other) are all above 0.8 for the PR curve. The AUC of Noise is low due to the high degrees of imbalance of noise data in the dataset.

4. Discussion

We have shown that the proposed algorithm using piecewise linear coefficients from the ECG beats is capable of detecting AF from ECGs recorded from wearable devices (AliveCor) with high accuracy (ranked among top 10 challenge results) by identifying important features of waveform morphology for AF. Detailed information regarding the PQRST waves helped improve the result significantly. One of the most significant challenges we faced during the development of our algorithm was the uncertainty about the clinical reasoning for the samples that were labeled as the 'Other' class. Hence, more information regarding the other arrhythmia-related diseases as well as further study about them could help improve the F_1 score. During the challenge, we tested other existing methods on the challenge data set, including wavelet entropy. We also tested our algorithm on other publicly available ECG datasets with 'AF' annotation, such as the MIT-BIH Arrhythmia Database (Goldberger *et al* 2000, Moody and Mark 2001). These findings are summarized in this section.

4.1. Other class

For our result, the 'Other' class has the lowest F_1 score, as shown in table 3. This is mostly because the 'Other' class includes any disease that is not AF. Our feature extraction primarily follows the diagnosis and clinical definition of AF. Some of the other arrhythmias include ventricular tachycardia, ventricular fibrillation, and supraventricular tachycardias which can be sub-classified based on atrial or AV nodal origin and AV nodal re-entry tachycardias (Wang and McClain 2017) (figure 9). Since our model is based on manually selecting features streamlined to detect AF without knowing the individual ECG characteristics of each specific type of arrhythmia in the 'Other' class, we have essentially limited the performance of the algorithm by over-fitting for the 'AF' class. For example, separation of noise and ventricular fibrillation was not investigated. Both of them are irregular rhythms and differentiating them tends to be a hard problem in pattern recognition.

**Table 4.** Comparison table of different methods.

Methods	F_1 score (%)				Specificity	Sensitivity	PPV
	Normal	AF	Other	Overall			
Wavelet transformation	0.87	0.70	0.58	0.72	0.981	0.601	0.753
Fourier transformation	0.90	0.63	0.66	0.73	0.983	0.570	0.770
Piecewise linear spline	0.90	0.80	0.72	0.81	0.982	0.779	0.812

4.2. Comparison to other methods

Wavelet transformation has been widely used in AF detection (Alcaraz *et al* 2006, Ródenas *et al* 2015, García *et al* 2016). We included the RR-interval information (distance between two R peaks and the difference between the lengths of heartbeats) and the coefficient and entropy of the wavelet transformation. This method achieved an overall F_1 score of 0.72 and accuracy of 0.74. We also replaced wavelet coefficients with Fourier coefficients as the features, which achieved a similar level of accuracy and F_1 score. Table 4 illustrates detailed results, revealing that the wavelet and Fourier coefficient methods have limited powers to detect AF on such datasets with the XGBoost classifier.

4.3. Other dataset

The method can easily be applied to other ECG datasets. We applied our model trained by the challenge data on a subset of the MIT-BIH Arrhythmia Database. For this dataset, the recordings are annotated as ‘AF’ or not AF (not ‘AF’ may include ‘Normal’, ‘Other’ or ‘Noisy’). The dataset includes 81 ECG samples, and their lengths are approximately 30 s. By using the built model from the challenge data, we achieved an accuracy of 93.83%, sensitivity of 87.5% and specificity of 95.89% with this dataset.

4.4. Limitation and future direction

Our algorithm extracts features at the heartbeat level, which can be computationally expensive when applied to long ECG recordings. Although we have used large computational resources, which may not be currently available for wearable devices, the data can be sent to a cloud environment for processing. Additionally, we did not have access to patients’ demographics, clinical history, or lab results. Such clinical information can give a better context for a higher accuracy of the AF classification.

Acknowledgement

Xiao Wang’s research is supported by NSF-DMS 1613060 and an i-GSDI award from Purdue University.

ORCID iDs

Vida Abedi  <https://orcid.org/0000-0001-7689-933X>

References

- Afdala A, Nuryani N and Nugroho A S 2017 Detection of atrial fibrillation using artificial neural network with power spectrum density of rr interval of electrocardiogram *J. Phys.: Conf. Ser.* **795** 012073
- Alcaraz R, Vayá C, Cervigón R, Sánchez C and Rieta J J 2006 Wavelet sample entropy: a new approach to predict termination of atrial fibrillation *Computers in Cardiology (IEEE)* pp 597–600
- Berg S M, Bittner E A and Zhao K H 2016 *Anesthesia Review: Blasting the Boards* (Philadelphia, PA: LippincottWilliams & Wilkins) pp 39–46

¹⁰ AV: atrioventricular.

- Camm A J *et al* 2010 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.* **31** 2369–429
- Canento F, Lourenço A, Silva H and Fred A 2013 Review and comparison of real time electrocardiogram segmentation algorithms for biometric applications *Proc. 6th Int. Conf. on Health Informatics*
- Carreiras C 2015 BioSPPy: biosignal processing in Python (<https://github.com/PIA-Group/BioSPPy>)
- Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. 22nd ACM Sigkdd Int. Conf. on Knowledge Discovery and Data Mining* (New York: ACM) pp 785–94
- Christov I I 2004 Real time electrocardiogram qrs detection using combined adaptive threshold *Biomed. Eng. Misc* **3** 28
- Clifford G, Liu C, Moody B, Lehman L, Silva I, Li Q, Johnson A and Mark R 2017 AF classification from a short single lead ECG recording: the PhysioNet Computing in Cardiology Challenge *Comput. Cardiol.* **44**
- Colloca R, Minardi L and Clifford G D 2013 Implementation and testing of atrial fibrillation detectors for a mobile phone application *Master's Thesis* Politecnico di Milano and University of Oxford
- Da Poian G, Liu C, Bernardini R, Rinaldo R and Clifford G D 2017 Atrial fibrillation detection on compressed sensed ECG *Physiol. Meas.* **38** 1405
- Daqrouq K, Alkhateeb A, Ajour M and Morfeq A 2014 Neural network and wavelet average framing percentage energy for atrial fibrillation classification *Comput. Methods Programs Biomed.* **113** 919–26
- Engelse W and Zeelenberg C 1979 A single scan algorithm for qrs-detection and feature extraction *Comput. Cardiol.* **6** 37–42
- Fuster V *et al* 2001 ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients with Atrial Fibrillation) developed in collaboration with the North American Society of Pacing and Electrophysiology *Circulation* **104** 2118–50
- Gamboa H 2008 Multi-modal behavioral biometrics based on HCI and electrophysiology *PhD Thesis* Universidade Técnica de Lisboa, Instituto Superior Técnico
- García M, Ródenas J, Alcaraz R and Rieta J J 2016 Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation *Comput. Methods Programs Biomed.* **131** 157–68
- Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet *Circulation* **101** e215–20
- Hamilton P 2002 Open source ECG analysis *Computers in Cardiology* (IEEE) pp 101–4
- Hnatkova K, Copie X, Staunton A and Malik M 1995 Numeric processing of lorenz plots of rr intervals from long-term ECG: comparison with time-domain measures of heart rate variability for risk stratification after myocardial infarction *J. Electrocardiol.* **28** 74–80
- Jones E *et al* 2001 SciPy: open source scientific tools for Python www.scipy.org/
- Jung Y and Kim H 2017 Detection of PVC by using a wavelet-based statistical ECG monitoring procedure *Biomed. Signal Process. Control* **36** 176–82
- Lip G *et al* 2016 Atrial fibrillation *Nat. Rev. Dis. Primers* **2** 1–26
- Mohebbi M and Ghasseminian H 2011 Prediction of paroxysmal atrial fibrillation using recurrence plot-based features of the rr-interval signal *Physiol. Meas.* **32** 1147
- Moody G B and Mark R G 2001 The impact of the MIT-BIH Arrhythmia Database *IEEE Eng. Med. Biol. Mag.* **20** 45–50
- Naccarelli G V, Varker H, Lin J and Schulman K L 2009 Increasing prevalence of atrial fibrillation and flutter in the United States *Am. J. Cardiol.* **104** 1534–9
- Nair T G, Geetha A and Asharani M 2013 Adaptive wavelet based identification and extraction of pqrst combination in randomly stretching ECG sequence *IEEE China Summit and Int. Conf. on Signal and Information Processing (ChinaSIP)* (Piscataway, NJ: IEEE) (<https://doi.org/10.1109/ChinaSIP.2013.6625344>)
- PhysioNet 2017 AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge www.physionet.org/challenge/2017/
- Rajpurkar P, Hannun A Y, Haghpanahi M, Bourn C and Ng A Y 2017 Cardiologist-level arrhythmia detection with convolutional neural networks (arXiv:1707.01836)
- Ródenas J, García M, Alcaraz R and Rieta J J 2015 Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms *Entropy* **17** 6179–99
- Sankari Z and Adeli H 2011 Heartsaver: a mobile cardiac monitoring system for auto-detection of atrial fibrillation, myocardial infarction, and atrio-ventricular block *Comput. Biol. Med.* **41** 211–20
- Savelieva I and Camm J 2008 Update on atrial fibrillation: part I *Clin. Cardiol.* **31** 55–62
- Tibshirani R J *et al* 2014 Adaptive piecewise polynomial estimation via trend filtering *Ann. Stat.* **42** 285–323
- Übeyli E D 2007 ECG beats classification using multiclass support vector machines with error correcting output codes *Digit. Signal Process.* **17** 675–84
- Übeyli E D 2009 Combining recurrent neural networks with eigenvector methods for classification of ECG beats *Digit. Signal Process.* **19** 320–9
- Wang X and Liu C 2018 Adaptive piecewise linear regression and deep learning manuscript, in preparation
- Zong W, Heldt T, Moody G and Mark R 2003 An open-source algorithm to detect onset of arterial blood pressure pulses *Computers in Cardiology* (IEEE) pp 259–62