Developing An Atrial Activity-Based Algorithm For Detection Of Atrial Fibrillation

Steven Ladavich, Behnaz Ghoraani*

Abstract— **In this study we propose a novel atrial activity-based method for atrial fibrillation (AF) identification that detects the absence of normal sinus rhythm (SR) P-waves from the surface ECG. The proposed algorithm extracts nine features from P-waves during SR and develops a statistical model to describe the distribution of the features. The Expectation-Maximization algorithm is applied to a training set to create a multivariate Gaussian Mixture Model (GMM) of the feature space. This model is used to identify P-wave absence (PWA) and, in turn, AF. An optional post-processing stage, which takes a majority vote of successive outputs, is applied to improve classier performance. The algorithm was tested on 20 records in the MIT-BIH Atrial Fibrillation Database. Classification combining seven beats showed a sensitivity of 99.28%, a specificity of 90.21%. The presented algorithm has a classification performance comparable to current Heartrate-based algorithms yet is rate-independent and capable of making an AF determination in a few beats.**

I. Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia. It affects an estimated 2.3 million United States citizens, and this number is only expected to increase as the general population ages. Automatic detection of AF could provide cardiologists with significant information for accurate and reliable diagnosis and monitoring of AF and is crucial for clinical therapy. However, monitoring AF remains an open area of research when the heart rate is controlled.

AF is the disorganized propagation of electrical activity in the atria that causes the atria to fail to contract in an organized fashion. As a result, the atrial depolarization wavefront, the P-wave, measured during sinus rhythm (SR) devolves into a series of fibrillatory waves (i.e., f-waves) in the surface electrocardiogram (ECG). Additionally, the fractionated electrical wavefront stimulates the atrioventricular (AV) node inconsistently, producing irregular ventricular contractions. Based on this observation, current methods to detect AF can be divided into two main approaches: (i) R-R interval (RRI) analyses detect the irregular ventricular heartbeat that is often associated with atrial arrhythmia, and (ii) atrial activity (AA) approaches detect a lack of organized atrial activity in the ECG. The most dominant methods in the literature [1]–[10] which boosts classifier performance for each of the two algorithms. are based on RRI analysis. However, current top-performing RRI-based algorithms process and make an AF determination using relatively long windows of data (50 to 100 beats), which can obscure the onset and offset of AF episodes. As a result, very short AF episodes may not be detected. Furthermore, RRI algorithms are rate-based and cannot work if the patient has a pacemaker or is taking rate-control drugs, or if other heart issues, such as atrioventricular (AV) block, occur simultaneously with AF.

Targeting an AA-based method allows for the rate-independent detection of very short episodes because determinations can be made with as low as a single beat. The main challenge with a PWA-based technique is that the ECG captures a nonstationary signal, and the atrial activity has a relatively low signal-to-noise ratio (SNR) compared to ventricular activity. Our proposed classifier is an anomaly detector that trains on normal SR P-waves and identifies P-wave absence as a beat which does not contain a P-wave similar to ones seen in the training set. As explained in Section II, the algorithm is executed in two phases: the training phase and the testing phase. In both phases, preprocessing and feature extraction is performed on the supposed P-wave. The training phase creates a statistical model that describes the feature space distribution of P-waves in the training set. The testing phase calculates a score that reflects the likelihood that a new beat contains a P-wave. In this way, P-wave absence (PWA) and, in turn, AF are detected. The proposed method is evaluated with a paroxysmal ECG database, and the results are reported in Section III.

II. Materials and Methods

A. Database and ECG Records

MIT-BIH AF Database from Physiobank [11] was used for validation of the proposed AF detection algorithms. The dataset includes 25 long-term (10 hr) ECG recordings with AF (23 paroxysmal and 2 persistent) and contains 299 AF episodes (about 93.4 h). Out of the total 23 paroxysmal recordings, this study uses only the 20 recordings that contain sufficient SR data to construct a training set.

B. Training Set Selection

A total of 10 minutes of training data for each record is selected from a period of SR at least 35 minutes long. Training intervals and the periods of wait time between the intervals are determined by considering what recordings are practical within the confines of a visit to one's clinician. Training data is selected from two 5-minute intervals spaced 10 minutes apart. Additionally, to ensure training data is not immediately following or preceding AF, a minimum of 7.5 minutes of SR were required to precede and succeed the two 5-minute training sets.

* corresponding author, Rochester Institute of Technology, Biomedical Engineering Department, Rochester, 160 Lomb Memorial Drive Rochester, NY 14623-5604, bghoraani@ieee.org

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

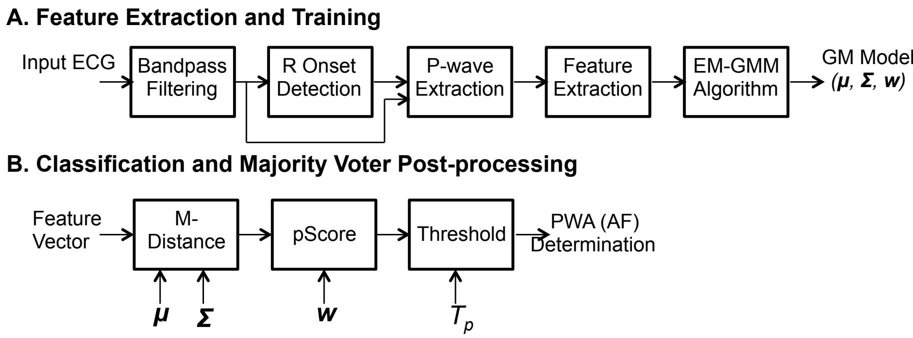

Fig. 1. Block diagram representing the proposed PWA algorithm.

C. R-wave Detection and P-wave Extraction

Figure 1 shows the proposed method's block diagram. A third-order butterworth bandpass filter is applied with poles at 0.5 Hz and 50 Hz to reduce baseline wander and powerline noise. Then R-wave detection and P-wave extraction are performed in two steps. First, R markers are placed at the point of maximum absolute derivative on the QRS complex. Then, P onset and P offset points are manually chosen from the segment preceding the R-wave marker, and supposed P-waves are extracted from the ECG recording.

D. Feature Extraction

A total of nine features are extracted: six features that estimate the morphology and three statistical features. The six morphological features are obtained by segmenting the P-wave into six contiguous sections and taking the mean value of each section. The three statistical features include variance, skewness, and kurtosis. Given an input vector, **p**, the variance is calculated as shown below:

$$V(\mathbf{p}) = \frac{1}{\ell_p - 1}\sum_{i=1}^{\ell_p}(\mathbf{p}_i - \mu_p)^2 \tag{1}$$

where V is the variance, ℓ_p is the length of the input vector, and μ_p is the mean of vector **p**. The skewness is calculated as follows:

$$S(\mathbf{p}) = \left(\frac{\sqrt{\ell_p(\ell_p-1)}}{\ell_p - 2}\right)\left(\frac{\ell_p^{-1}\sum_{i=1}^{\ell_p}(\mathbf{p}_i - \mu_p)^3)}{\left(\ell_p^{-1}\sum_{i=1}^{\ell_p}(\mathbf{p}_i - \mu_p)^2\right)^{3/2}}\right) \tag{2}$$

where S is the skewness, ℓ_p is the length of the input vector, and μ_p is the mean of vector **p**. The kurtosis is calculated as given below:

$$K(\mathbf{p}) = \left(\frac{\ell_p - 1}{(\ell_p-2)(\ell_p-3)}\right) \times \left((\ell_p+1)\left(\frac{\ell_p^{-1}\sum_{i=1}^{\ell_p}(\mathbf{p}_i - \mu_p)^4)}{(\ell_p^{-1}\sum_{i=1}^{\ell_p}(\mathbf{p}_i - \mu_p)^2)^2}\right) - 3(\ell_p - 1)\right) + 3 \tag{3}$$

where K is the kurtosis, ℓ_p is the length of the input vector, and μ_p is the mean of vector **p**. All the statistical features are calculated as sample statistics and additional terms to adjust for bias are included.

E. Training

As shown in Figure 1A, training begins by extracting a feature vector for each beat in the training set. The feature space distribution of SR P-waves is used to create a statistical model that generalizes SR P-waves. The nine-dimensional feature space is modeled by a multivariate Gaussian mixture model (GMM). The Expectation-Maximization (EM) algorithm is employed to generate the model for the training set. Because the EM algorithm operates with a predefined number of hidden distributions and its outcomes vary for different initializations, the optimal model is determined iteratively. Beginning with one hidden distribution, ten models are created. The model with the most successful initialization is chosen. Similarly, a best model with two hidden distributions is determined. The number of hidden distributions is increased until the quality of the modeling (measured by log-likelihood) ceases to significantly improve by more than 1%.

The iterative application of the EM algorithm converges on a model to describe the feature space distribution of the training set. The model is defined by the set of unique parameters μ, $\mathbf{\Sigma}$, and $\boldsymbol{w}$ for each hidden distribution. These parameters are used in the test/processing stage to detect AF.

F. Classification

Test set evaluation begins with P-wave feature extraction as shown in Figure 1B. The classifier determines if the feature vector constitutes PWA or AF. The first classifier stage calculates the Mahalanobis distance of the feature vector from each hidden distribution center. The calculation is shown in Equation 4:

$$M_j(f) = \sqrt{(f-\mu_j)^T \mathbf{\Sigma_j}^{-1}(f-\mu_j)} \tag{4}$$

where f is the feature vector, j is the index of the hidden distribution, M_j is the M distance with respect to j^{th} hidden distribution, μ_j is the mean/center of the j^{th} hidden distribution, $\mathbf{\Sigma}_j$ is the covariance matrix of the j^{th} hidden distribution and T denotes the matrix transpose.

The M distance is then scaled by a spread parameter which adjusts the generalization of the model. Increasing the spread allows for a wider range of variation in normal SR. Adjusting the spread parameter has the same effect as adjusting the threshold parameter although the relationship between the two is inverse and non-linear. For the purposes of this method, the spread was empirically chosen with a value of 500. Smaller or larger values result in less fluent optimal thresholds, which must be specified with greater precision much closer to 0 or 1, respectively. Each scaled M distance is passed through a radial basis kernel function, scaled again according to the prevalence of that distribution in the training set, and summed together. This described calculation is termed the *pScore* and is shown in Equation 5:

$$pScore(f) = \sum_{j=1}^{N_Z} w_j \ exp(-(\frac{ln(\frac{1}{2})}{s})M_j^2) \tag{5}$$

where N_Z is the number of hidden distributions, s is the spread parameter (i.e., $s = 500$), w_j is the weight applied to the j^{th} distribution and is equal to the probability of the j^{th} distribution.

The *pScore* reflects the likelihood that a feature vector extracted from the segment preceding the QRS complex is also a P-wave. A *pScore* of 1 reflects a near absolute certainty that the evaluated segment is a P-wave, while a *pScore* of 0 reflects a near absolute certainty that the evaluated segment is not a P-wave. The *pScore* is compared to an empirically determined threshold, T_p (as will be explained in Section III), to determine whether or not PWA and, in turn, AF are present.

G. Majority Voter Post-processing

The final stage of classification is the implementation of the optional majority voter post-processing stage. Cases of noise, artifact, and/or ectopy isolated to a single beat that potentially cause classification errors and are not clinically relevant can be effectively eliminated. By taking a majority vote of an odd number of sequential AF determination outputs, classification errors in a minority of beats are ignored.

III. Results

The proposed method was applied to the AF database in Section II-A. The classifier proposed in this work generates statistical models specific to each record by training with patient-specific training sets. The average number of hidden distributions that the iterative EM algorithm converges upon

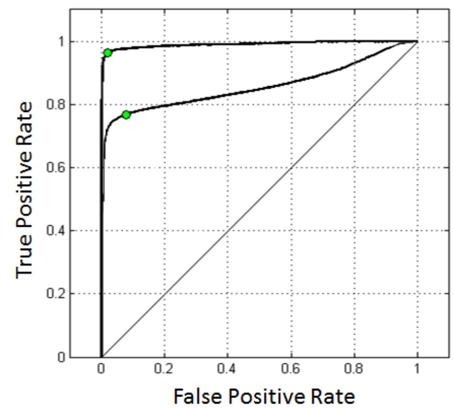

Fig. 2. ROC curves for the best performing record (Record 04936) and the worst performing record (Record 06995) with area under the curve of 0.99 and 0.86, respectively. The circles on the graph locate the optimum threshold for each patient.

TABLE I

Table of Optimum Threshold Value for Each Record

ID	Optimum Threshhold	ID	Optimum Threshhold
04015	0.74	04043	0.55
04048	0.69	04126	0.88
04746	0.83	04908	0.65
04936	0.75	05091	0.84
05121	0.46	05261	0.83
06453	0.74	06995	0.67
07879	0.83	07910	0.88
08215	0.69	08219	0.65
08378	0.55	08405	0.70
08434	0.83	08455	0.75

is 3.75 hidden distributions and ranges between 2 and 5 hidden distributions. This result confirms that the statistical distribution of the P-wave feature distribution is relatively complex and warrants being modeled by the GMM used in this study. We also found that the correlation of the number of training beats and the number of hidden distributions is equal to 0.20. The low correlation between training beats and the number of hidden distributions suggests training set sizes are sufficiently large.

Processing features extracted from each beat results in the unthresholded classifier output, the *pScore*, which is a reflection of the likelihood that a particular beat contains a P-wave. The *pScore* is compared to a threshold to acquire the final classification. Varying this threshold and calculating the sensitivity and specificity generates the receiver operating curve (ROC), which can be used to empirically determine the optimal threshold value. ROC curves and the optimal threshold values are generated automatically using MATLAB's *perfcurve* function. Patient-specific ROC curves for the best and worst cases along with the location of the optimum threshold are presented in Figure 2. These optimal threshold values are listed in Table I.

Using the optimal thresholds determined from the ROC curves (Table I), the test set is processed for each record. Furthermore, the test set is processed with majority voter

	Mean				Median			
Vote Size	Se (%)	Sp (%)	PPV (%)	Err (%)	Se (%)	Sp (%)	PPV (%)	Err (%)
1	90.84	86.78	59.62	11.78	91.98	86.18	70.70	13.02
3	95.76	88.89	62.56	8.77	97.45	88.01	76.02	9.40
5	96.70	90.46	64.40	7.42	98.67	90.04	80.39	7.80
7	97.23	90.11	64.49	7.14	99.28	90.21	80.42	7.12

TABLE II

MEAN AND MEDIAN PERFORMANCE RESULTS WITH VARIOUS MAJORITY VOTER SIZES AND THRESHOLDS SET TO THE OPTIMAL THRESHOLD VALUES.

post-processing described in Section II-G. Results in which the voting consists of 1,3,5, and 7 beats are presented in Table II. Mean and median results across all 20 records are presented for both threshold applications and for all 4 majority voter settings. Several observations can be made from the classifier performance results in table II: (i) Increasing the number of beats included in the majority voting improves classifier performance, but performance gains decrease with each additional included beat. (ii) Generally, the median classifier performance is higher than the mean classifier performance, indicating negative skewness. Negative skewness for each threshold setting (median >mean) corresponds to a majority of records producing better results with a few poor performing outliers.

The proposed algorithm is compared against other algorithms using statistical classification measurements including sensitivity (Se), specificity (Sp), and the length of the window (WL) that is evaluated to determine AF (see Table III). Many of the algorithms ([1]–[7]) were selected for comparison because they were identified by Larburu *et al.* [8] as the best performing results for particular methods; further evaluation of the algorithms was also provided by Larburu *et al.*. Additional algorithms selected for comparison represent more recent work, including ectopic beat removal ([9], [10]) which boosts classifier performance for each of the two algorithms. Our algorithm is capable of classifying AF with a very high sensitivity (99.28%), which is the highest among all compared algorithms with a comparable specificity. It is capable of making a rate-independent AF determination with fewer beats and a smaller window length than other algorithms. Due to the finer temporal resolution of the proposed algorithm, shorter episodes can be detected and AF onset and offset can be identified more accurately.

Algorithm	Method	WL (seconds)	Se (%)	Sp (%)
Moody et al. 1983	RRI	60	87.54	95.14
Slocum et al. 1992	PWA/FSA	180	62.80	77.46
Cerutti et al. 1997	RRI	90	96.10	81.55
Tatento et al.2001	RRI	50	91.20	96.08
Logan et al. 2005	RRI	120	87.30	90.31
Couceiro et al.2008	RRI + PWA/FSA	60	96.58	82.66
Babaezaideh et al. 2009	RRI/FSA	40	87.27	95.47
Dash et al. 2009	RRI	128 beats	94.4	95.1
Huang et al. 2011	RRI	101 beats	96.1	98.1
Proposed Work 2014	PWA	7 beats	99.28	90.21

TABLE III

COMPARISON OF AF DETECTION ALGORITHMS INCLUDING ADDITIONAL EVALUATION PERFORMED BY LARBURU *at al.* [8]. UNREPORTED PERFORMANCE MEASURES ARE DISPLAYED AS A DASH.

IV. CONCLUSION

In this study, we proposed a novel feature extraction and classification approach that detects AF through direct analysis of AA. We derive nine statistical and morphological features from SR P-waves and train a multivariate GMM-based classifier to model the distribution of SR P-waves in feature space. We then detect any P-wave variations that occur during AF, thereby detecting AF episodes by determining the absence of SR P-waves in the ECG recordings. The proposed algorithm achieves performance measures comparable to RRI-based algorithms despite only using a few beats (seven beats), whereas the majority of RRI algorithms utilize 50 to 100 beats. The single-beat, rate-independent nature of the proposed algorithm could provide clinicians with a reliable tool to detect the SR and AF transition that will happen over a few beats with no limitations for patients with rate-controlled drugs or pacemakers.

REFERENCES

[1] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, "Improvements in atrial fibrillation detection for real-time monitoring," *J. Electrocardiogr*, vol. 42, no. 6, pp. 522–526, 2009.

[2] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi, "Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes," *Computers in Cardiology*, vol. 24, pp. 77–80, 1997.

[3] R. Couceiro, P. Carvalho, M. Antunes, M. Harris, and J. Habetha, "Detection of atrial fibrillation using model-based ECG analysis," in *International Conference on Pattern Recognition*, 2008.

[4] B. T. Logan and J. Healey, "Robust detection of atrial fibrillation for a long term telemonitoring system," *Comp in Card*, vol. 32, pp. 619–622, 2005.

[5] G. B. Moody and R. G. Mark, "A new method for detecting atrial fibrillation using R-R intervals," *Comp in Card*, 1983.

[6] J. Slocum, A. Sahakian, and S. Swiryn, "Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity," *Journal of Electrocardiography*, vol. 25, no. 1, pp. 1–8, 1992.

[7] K. Tatento and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals," *Medical & Biological Engineering & Computing*, vol. 39, no. 6, pp. 664–671, 2001.

[8] N. Larburu, T. Lopetegi, and I. Romero, "Comparative study of algorithms for atrial fibrillation detection," *Comp in Card*, 2011.

[9] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, "Automatic real time detection of atrial fibrillation," *Annals of Biomed Eng*, 2009.

[10] C. Huang, S. Ye, H. Chen, D. Li, F. He, and Y. Tu, "A novel method for detection of the transition between atrial fibrillation and sinus rhythm," *IEEE TBME*, vol. 58, no. 4, pp. 1113–1119, 2011.

[11] A. L. Goldberger, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).