Detection of Atrial Fibrillation in ECGs

Tracy Chou, Yuriko Tamura, and Ian Wong {TYCHOU, YTAMURA, IANW}@STANFORD.EDU

1. Overview

Automatic detection and classification of arrhythmia in electrocardiograms (ECG) provides a framework for efficient diagnosis and broader outreach to patients at risk for cardiac diseases. While prevalent types of arrhythmia include premature ventricular contractions (PVC) and atrial fibrillation, the majority of existing literature focuses on automatic detection and classification of the former type. In this report, we discuss our heuristic and implementation for detection of atrial fibrillation, using a dataset provided by the MIT/BIH Arrhythmia Database. Using features from Fourier analysis, wavelet transformation, and R-R interval analysis, linear discriminant analysis (LDA) on individual segments performed well with classification error of approximately 10%. Our detector, which built on top of our classifier, successfully identified regions of atrial fibrillation with less than 2% error.

2. Motivation

2.1. Background

Electrocardiograms (ECGs) are recordings of the heart's electrical activity and are widely used by physicians to diagnose pathologies related to the heart. Normal (sinus) rhythms manifest as periodic time signals representing a series of heart beats, each with characteristic peaks that correspond to events during a single heart beat. Patients with or at risk of cardiovascular diseases often present ECGs that are irregular in rate and in morphology of the signal.

There is a recognized industry in creating automatic detection algorithms for arrhythmia, because it is impractical for doctors to comb through ECG data by hand. Episodes of arrhythmia may be infrequent, and recordings of more than 48 hours are often necessary to catch them.

However, while prevalent types of arrhythmia include both premature ventricular contractions (PVC) and atrial fibrillation, most prior research has focused on PVC. The techniques used for detecting PVC, which occurs as singular beats, cannot be applied directly to detecting atrial fibrillation, which takes place over a sequence of heartbeats and can be observed by irregularity of both morphology and rhythm of heartbeats.

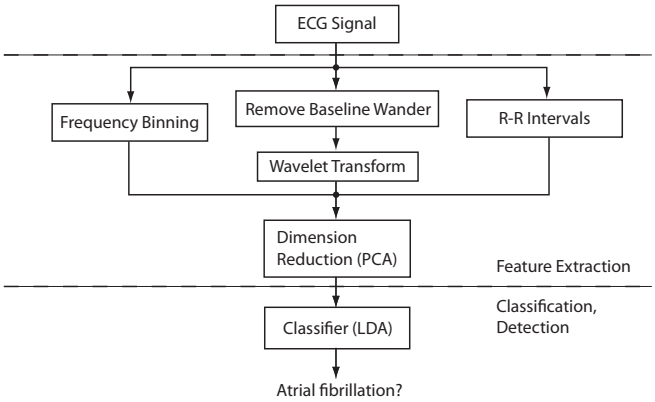

Figure 1. Our overall approach.

2.2. Problem Statement

Given an entire ECG recording, can we detect regions with atrial fibrillation?

3. Our Approach

Our approach to addressing the problem of atrial fibrillation detection is depicted in Figure 1. Using the MIT-BIH Arrhythmia database as a signals database, we first trained our binary classifier, which used linear discriminant analysis, on both normal and atrial fibrillation regions in these signals. The features we used are detailed in a Section 5. Afterwards, we applied our binary classifier to regions of our test signal. More precisely, we classified rolling regions over our test signal, assigning all points in the signal a probability score for atrial fibrillation. For verification we compared our predicted regions of atrial fibrillation to regions of actual atrial fibrillation as annotated by the database.

4. Obtaining ECG Data

PhysioNet provides access to various ECG datasets, including the MIT-BIH Arrhythmia Database, which provides beat and rhythm annotations manually done

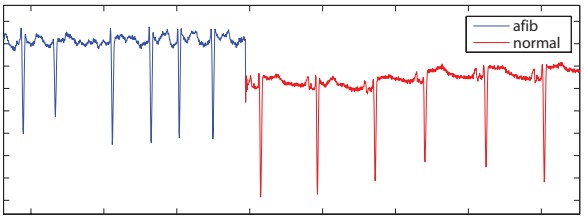

Figure 2. Segments of atrial fibrillation and normal beats.

by physicians. The Waveform Database (WFDB) Library provides C functions to decode the data and annotations, which we were able to port into MATLAB. The database is comprised of 48 fully annotated half-hour, two-lead ECG recordings. Of these, 7 recordings contain atrial fibrillation – these 7 were our signals of interest.

5. Feature Selection

We considered several different features to use in classification of individual heartbeats. The most successful set of features was a combination of all the features we considered - frequency components, time series data, and length of the beats.

5.1. Defining Training Examples

One straightforward approach would have been to use individual heartbeats as training examples, since the end goal was to classify individual beats. However, consecutive beats tend to be very similar, and this approach would have given more weight to training examples that occur in long continuous runs. Especially given the few datasets available, this could have skewed the classification algorithm considerably. Our approach therefore segmented the records according to the annotations of "normal" or "atrial fibrillation", and each segment was considered one training example. Each training example could have a variable number of beats.

We only used records containing atrial fibrillation; we did not use normal segments from any records not containing atrial fibrillation.

5.2. Frequency Components

We selected frequency features by applying the Fourier transform on the raw data. We binned the contributions in bins of 10 Hertz; that is, the frequency components between 0 and 10 Hertz would be summed, between 10 and 20, and so on. We then computed the power (in decibels) by taking the log of these sums, and normalized them. We experimented with different

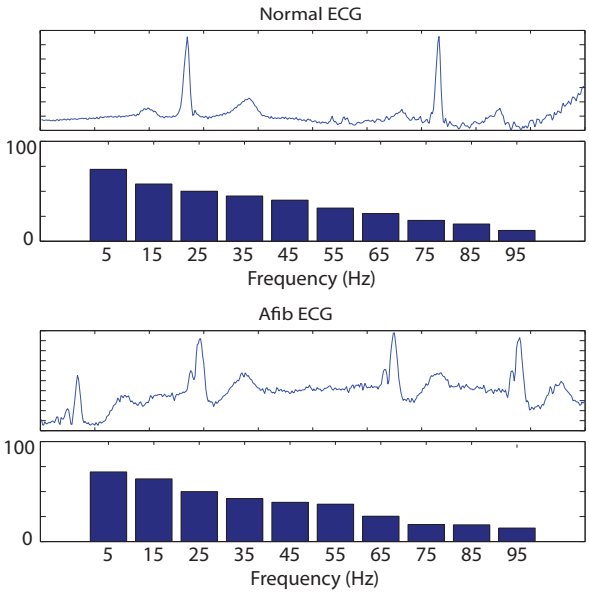

Figure 3. Top: Normal signal and histogram of frequencies after Fourier transform. Bottom: Same, but for signal with atrial fibrillation.

size bins, and different numbers of bins (which corresponded to number of features). In general we found that bins of 10 Hertz were best, and also that the frequencies above 200 Hertz reflected noise in the ECG signals and were not representative of wave morphology or timing.

5.3. Time Series Data

We selected time series features by performing wavelet transform, windowing individual beats, downsampling those beats, and averaging the samples for each window.

ECG recordings frequently exhibit baseline wander – artificial, fluctuating curves that offset entire ECG signals – due to various sources of recording noise, including patient movement and mechanical displacement of the ECG leads. Before extracting time series features, we first preprocessed the signal to remove baseline wander. We compared different ways of filtering the signal to recover the baseline, including single-median, double-median, double-mean, and lowpass filters. We found that a double median filter, first with width 300 milliseconds and second with width 600 milliseconds, was most effective for retrieving the baseline, which we then subtracted from the original signal.

The next step was to smooth out noise and enhance morphological features of the signal. Based on results in literature, particularly from Andreao *et al.*, and our own experimentation, we chose to use the wavelet

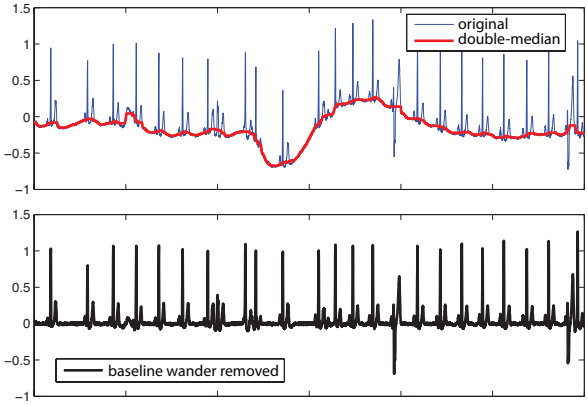

Figure 4. Top: ECG with baseline wander and the signal after double-median filter. Bottom: ECG with baseline wander removed.

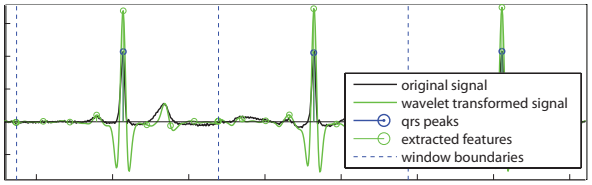

Figure 5. Wavelet transformed signal with extraction windows and selected features.

transform with the "Mexican hat" wavelet, given by $\mathcal{W}f(t,s) = f * \bar{\phi}_s(t) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t)\phi^* \left(\frac{\tau - t}{s}\right) d\tau$. Andreoa *et al.* [Andreao07] compared various wavelet transform, and showed that the Mexican hat wavelet is similar in shape to ECG signals and is the ideal candidate to enhance the major features of the signal while removing high frequency noise. After testing different scales of the wavelet transform on our ECG signals, we settled on the scale $s = 2^2$.

To determine the locations of individual beats, we used the MIT database annotations for QRS peaks and windowed by splitting beats halfway between peaks. After windowing, we anchored the QRS peak of each window as the halfway point – as it was the most significant feature – and uniformly sampled points before and after the peak to get the desired number of features.Our first approach was to find local maxima and minima within subintervals of the window, but the more straightforward downsampling produced better results in classification.

Finally, for each training example, we averaged the features obtained per beat in the example.

5.4. Beat Length

Because our windows were of variable size, we used the length of windows as a feature. In fact, much previous work on analyzing ECG signals considered R-R intervals as an important characteristic in differentiating normal and abnormal beats. However, we found it easier to substitute length per beat as an approximate measure of R-R intervals, which would have required looking at previous and subsequent beats. We grouped this feature with the time series data in our testing.

5.5. Combining Feature Sets

After testing the effectiveness of the classifier using the different feature sets, we found that combining all of them gave the best results.

For the frequency and time series analysis, the number of features extracted per training example was very straightforward, corresponding to frequency bins and downsampled points, respectively. When testing the combination of sets, for a desired number of features n, we concatenated n features for each set to get $2n$ features. However, when concatenated, we found our training matrix of a higher dimension to be not full rank, limiting our inference abilities. That is, the higher dimensional data actually resides in a lower one, so we applied principal component analysis to reduce the dimensionality back down to n. On the downside, PCA made it difficult to attach physiological meaning to the features we extracted.

6. Classification

Our goal for classification was binary classification of segments of ECG signals as normal beats or atrial fibrillation.

6.1. Classification Algorithm

We used standard Gaussian discriminant analysis with a pooled covariance matrix as our classification algorithm. We also tried Gaussian discriminant analysis with covariance matrices stratified by group, as well as Naive Bayes, with less success.

6.2. Performance Metrics

We used leave-one-out cross-validation to evaluate the performance of our classifier, since we were working with a limited dataset. The best performance was 9.17% error using 10 combined features, with 92.31% sensitivity and 90% specificity. The lowest classification error using only frequency features was 12.84%, with 20 features, measuring frequency components up to 200 Hertz. The lowest classification error using only temporal components was 18.37%, with 10 features (9 samples per window, combined with beat length).

Classifying with only the individual frequency or temporal feature sets had perfect sensitivity but no specificity, indicating that the classifier always erred on the side of predicting atrial fibrillation.

# features	Classification error		
	Fourier	Wavelet + RR	Both
5	0.2569	0.2081	**0.1193**
10	0.1927	0.1837	**0.0917**
15	0.1651	0.2110	**0.0917**
20	0.1284	0.2569	**0.1101**
25	0.1651	0.2110	**0.1284**
# features	Sensitivity = # TP / (# TP + # FN)		
	Fourier	Wavelet + RR	Both
5	1	1	**0.8974**
10	1	1	**0.9231**
15	1	1	**0.8718**
20	1	1	**0.8974**
25	1	1	**0.8205**
# features	Specificity = # TN / (# TN + # FP)		
	Fourier	Wavelet + RR	Both
5	0	0	**0.8714**
10	0	0	**0.9000**
15	0	0	**0.9286**
20	0	0	**0.8857**
25	0	0	**0.9000**

7. Detection

An automatic detection tool for atrial fibrillation is the primary application of our rhythm classification tool – the original motivation of our problem was that arrhythmia are often only captured in Holter recordings that are too long for visual scanning by doctors. Using our best performing set of features for classification, we developed an algorithm to detect episodes of atrial fibrillation in a test ECG record, with only 1.75% error on our test data.

To classify sections of a 30-minute MIT database record, we considered a test window which we slid through the record with constant increments on the starting index. From each window we extracted 10 features combining Fourier transforms, wavelet transforms, and R-R interval analysis, and classified the segment with linear discriminant analysis, as described above. Test windows classified as atrial fibrillation were output as an array of 1's for the represented sample indices, and 0's otherwise. The relative probability of finding an arrhythmia at a particular index was calculated as the normalized cumulative sum of all outputs of test windows that included the point. If this probability was greater than 0.5, we predicted arrhythmia, and normal otherwise.

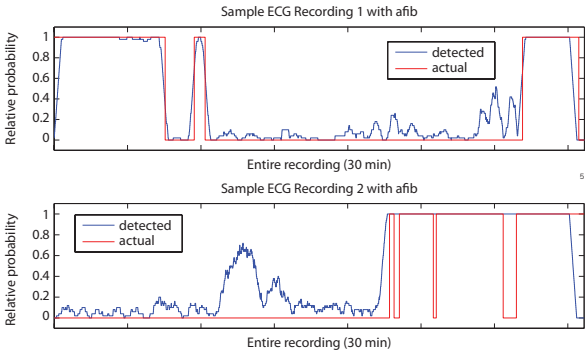

Figure 6. Our detector successfully identified regions with atrial fibrillation.

We used a constant length of 5000 samples per test window, corresponding to 25 seconds. The offset between each start index was set to be 100 samples, corresponding to 0.5 seconds. We chose these numbers because segments of arrhythmia often last several minutes, and 0.5 seconds corresponds to approximately half a beat in a normal sinus rhythm and up to a whole beat for irregular rhythms. The relative fraction of the offset to the test window length defines the vertical resolution of the additive algorithm, while the absolute value of the offset changes horizontal resolution.

8. Conclusions

Machine learning applied to ECG analysis provides a platform for accurate and efficient arrhythmia diagnosis without extensive knowledge of the mechanisms or characteristics of different classes. Through experimentation with different feature extraction and machine learning techniques, we developed a robust binary classification method for atrial fibrillation. Detection using our wave classifier shows promise in clinical application of our algorithm.

Our approach first represented different heartbeats with characteristic features combining frequency domain analysis as well as time series data, and then used these features to train a linear discriminant analysis classifier. The standalone classifier had approximately 90% accuracy, but our detection tool performed even better, successfully detecting regions of atrial fibrillation with only 1.75% error.

Further work to improve the performance of our classification and detection algorithms should incorporate different ways of dividing up segments for feature analysis. In our temporal analysis we considered beats individually, but one researcher, Omer Inan, has suggested that larger windows of 3 beats are preferable,

because they capture rhythm transitions. Our results showed that combining time series data with Fourier transformed data drastically improved performance, indicating the importance of longer scale wave characteristics such as heart rate for classification of atrial fibrillation.

9. Acknowledgments

We would like to thank Jinglin Zeng, Ph.D., Uday Kumar, M.D., Mark Day, Ph.D., and Omer Inan for providing us with inspiration, expertise, and moral support throughout the project. (We would also like to designate The Killers' song "Human" as the official theme song of our project.)

10. References

Andreao RV, Dorizzi B, Boudy J. ECG Signal Analysis Through Hidden Markov Models. IEEE Transactions on Biomedical Engineering, 53(8):1541-1549, 2006.

Andreao RV, Boudy J. Combining Wavelet Transforms and Hidden Markov Models for ECG Segmentation. EURASIP Journal on Advances in Signal Processing, 10.1155/56215:8pp, 2007.

Coast DA, Stern RM, Cano GG, Briller SA. An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models. IEEE Transactions on Biomedical Engineering, 37(9):826-836, 1990.

Hughes NP, Tarassenko L, Roberts SJ. Markov Models for Automated ECG Interval Analysis.

Koski A. Modelling ECG signals with hidden Markov models. Artificial Intelligence in Medicine, 8:453-471, 1996.

Laguna P, Mark RG, Goldberg A, Moody GB. A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. Computers in Cardiology, 24:673-676, 2007.

Inan O, Giovangrandi L, Kovacs G. Robust Neural-Network-Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features. IEEE Transactions on Biomedical Engineering, 53(12):2507-2515, 2006.