

Classification of Electrocardiogram Anomalies

Karthik Ravichandran, David Chiasson, Kunle Oyedele
Stanford University

Abstract

Electrocardiography time series are a popular non-invasive tool used to classify healthy cardiac activity and various anomalies such as Myocardial Infarction, Cardiomyopathy, Bundle branch block, myocardial hypertrophy and Arrhythmia. While most ECGs are still read manually, machine learning techniques are quickly being applied for diagnosis due to the advantages of being quick, cheap, and automatic as well as the for the ability to examine features not obvious to the human eye. In this paper, a binary classification algorithm is implemented, and it reports a classification accuracy of **97.2%**. The heart beats are analyzed, and a prediction is made to mark the beat as normal/ abnormal. Four major classes of features are - morphological, global trend prediction, regularity, and beat energy. The classifier is based on Support Vector Machine algorithm, and learning is optimized by feature selection (Random Forest), dimensionality reduction (PCA), and Kernel optimization.

1 ECG Overview

An electrocardiograph (also known as EKG or ECG) is a graph showing the electrical impulses of the heart over time. A single pulse consists of three major section called the P-wave, the QRS complex, and the T-wave as shown in Figure [1].

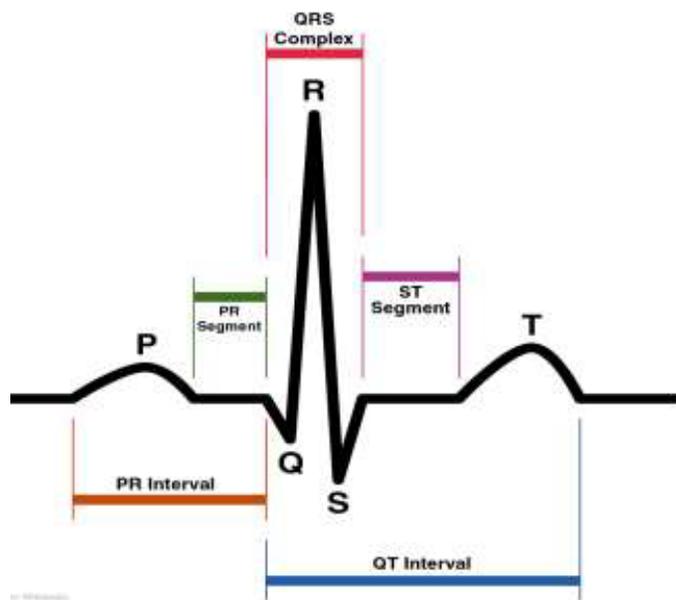


Figure 1: A single EKG pulse example [1]

Based on conversations with health care professionals, there are several aspects of the ECG trace which are particularly helpful in diagnosing anomalies. These consist of Morphology (Frequency Response), Heart Rate (Bpm), Regularity (and second moment), existence of wave segments and relative amplitudes(P,QRS,T), Timing Intervals(P-Q, P-R, QRS, S-T), normalized energy in a a beat. Evidently, various segments of the beat correspond to pumping action of various chambers of the heart, and in a way, the abnormality can be localized.

Medical professionals adopt a systematic approach to interpreting an ECG, which consists of observing various aspects of the ECG trace and asking certain questions. The steps are [2]:

- Analyze the rate of atria and ventricles - is the rate between 60 and 100?
- Analyze the rhythm - is the rhythm regular or irregular, and is it regularly irregular or irregularly irregular?
- Analyze the axis - is there a left or right axis deviation?
- Analyze the P-R interval - is the interval between normal parameters and length (i.e. less than 0.2 sec)?
- Analyze the P-waves - are there P-waves? Is there a P-wave for every QRS complex; is there a QRS complex for every P-wave?
- Analyze the QRS complex - is it wide or narrow? Is there a Q wave? is there a delta wave?
- Analyze the ST segment-T wave Is there ST elevation or depression? Are the T waves inverted?
- Overall interpretation

2 Data Source

The data used to train our algorithm and perform tests were obtained from the venerable MIT-BIH Arrhythmia Database [3] with 48 half-hour excerpts of two-channel ambulatory ECG recordings. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two more cardiologists independently annotated the beats and classified into one of 23 cases. The distribution of beats is that **66.67%** of the beats are “Normal”. Here, we just deal with two classes (Normal or not), and it is not difficult to extend this project to a multi-category classification problem.

3 Preprocessing

Each beat and its label were extracted from individual patient records, and our database consisted of 112548 beats. The beats are then temporally normalized, to ease feature extraction. Each beat now is interpolated or decimated to 600 samples, and this number corresponds to a 36 beats per minute heart rate (@ 360 Hz). Several beat information such as Maximum, minimum, mean, median amplitudes, instances of (R) peaks of the current beat and its neighbors. Before normalizing temporally, the beat duration is stored as well. The data is then normalized in amplitude so as to avoid amplitude variations affect the features. Various features are calculated in a local window (say 50 beats on either side), and local amplitudes such as local mean, and median are computed to predict the global trend of the heart activity. Unorthodox beats are handled appropriately.

4 Feature Extraction

Our four classes of features were intuitively reasoned and are independent of amplifier gain, filter response, and other hardware dependencies.

(a) Morphology

These set of features quantify the shape and structure of a beat.

i Wavelets

Wavelet coefficients are used to represent the overall morphology of a beat. We used MATLAB’s wavelet decomposition toolbox, and observed that the number of coefficients (maximum) required to faithfully define a beat was 50 out of 660 coefficients. This decomposition filters the signal by removing high frequencies. A debauchies-4 wavelet was used for a 6-level decomposition. In fact, Maximum two numbers can satisfactorily describe the signal as in Figure [2]. Wavelet coefficients are computed by correlating the wavelet at every time instance, and varying the scale of the wavelet at all points. These measurements do not occur at Zero bandwidth like in a Fourier transform. Wavelets exploit the uncertainty in time-frequency, and each measurement provide something from both the temporal and frequency domains, to satisfactory accuracy.

ii Beat segment characteristics

Four features are extracted here to depict the relative properties of P wave, QRS complex, and T wave. P wave characterize the pumping activity of the atrium, QRS- the ventricles, and T wave - the re-polarization of the ventricles. These create a cycle, and the absence or malfunction of one can seriously affect the circadian rhythm. So, the features are amplitude ratio of QRS/P, QRS/T, temporal (duration) ratio of QRS/P, and QRS/T.

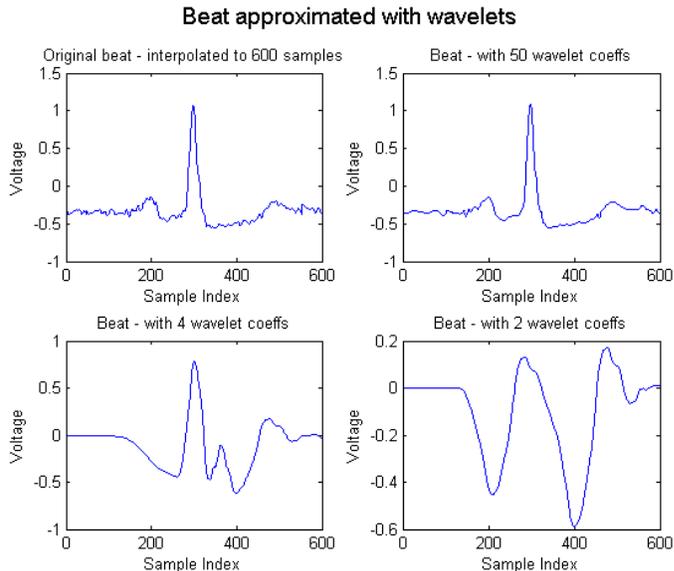


Figure 2: Wavelet Approximation

iii P-R interval

PR interval quantifies the electrical conduction of the heart from atrium to ventricles.

iv Instantaneous Bpm

At every beat, we calculate the instantaneous bpm (beats per minute), by computing the ratio $2/T$, where T is the distance between previous and next beats.

(b) Energy in a beat

To characterize the pumping power of various segments, the normalized areas of P segment (Pre-peak), QRS complex (peak), and T segment (Post-peak) are computed. The points of splitting is chosen by finding a weighted mean of the maximum and the mean points in a beat. (Figure [3]).

(c) Regularity

This is one of the novel features of this work. When the temporal instances of a maxima of a periodic waveform (like that of a healthy heart) is fitted against a monotonically running uniform index, we get a linear fit. Here, we perform linear regression to fit a line to the peak instances in the entire patient record, and calculate the “Residues” (outliers) for every beat’s peak.

This can quantify the regularity of the heart beats for a record, and a second moment can predict if the beats are regularly irregular, or irregularly irregular. We see two other variants of the features in the next section on global trends.

A small disadvantage of this method is that, we assume that holistically healthiness of the record. Global rhythm (not subtle) will be considered separately as a global feature, so, overlooking this fact is tolerable. A +ve residue

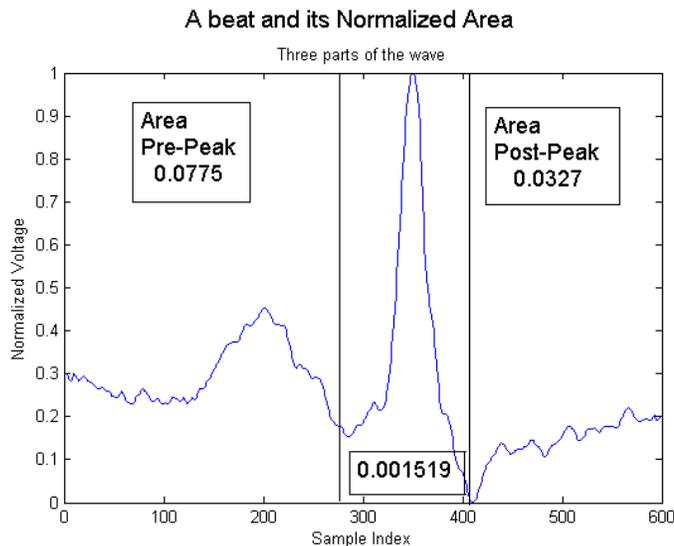


Figure 3: A beat with its Norm Areas

characterizes a slower beat, while a -ve residue implies a faster beat.

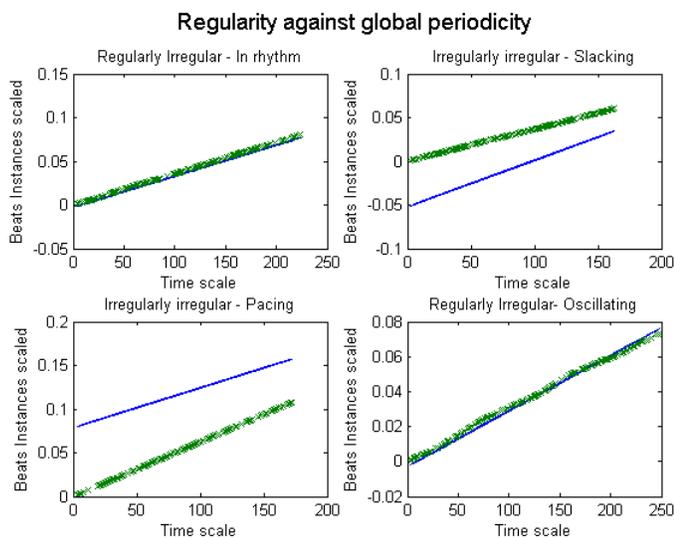


Figure 4: Regularity and Residues

(d) Global Trend

It is inevitable to include features, besides that of beats, that tend to predict the global trend of the record. It is, of course, not useful to incorporate a global feature here, as it gets redundant when we have too many beats from a single record. Two such trend predictors are computed from the regularity residues. A local residue measure is calculated by summing over residues over a local window. “Regularly irregular” rhythm has near zero value as the +ve, and -ve residues tend to cancel out, whereas, sum of many +ve, or many -ve depicts a potential “irregular irregularity”. The second - feature is the sum of squares of residues over a local window. This can quantify the irregularity in that window. Apparently, this feature has

a near zero importance, as will be seen from the feature selection routine.

Other predictors that are used are Variance in beat duration, with respect to the global mean of the record. This characterizes the regularity of the entire beat, and here too, we assume a healthy bpm globally. Other ratios such as (min/max), (max/Local mean), and (max/Local median) are calculated as part of morphology with respect to a local window.

(e) Global Features

Although we don’t incorporate global features in this work, any useful decision about the record can be made by utilizing other global features apart from the beat labels. We observed few of these to provide information such as health of global rhythm, beat duration, and bpm, FFT of beat temporal behavior, and distribution of each class of beats.

5 Classification

The following sections report our flow to optimize the learning process. The steps can be in any order. Obviously, this flow could be iterated until necessary performance is attained.

(a) Data Sampling

We sampled 10,000 beats (arbitrary, and computationally easy) from our database of 112548 beats. The sampling has to be uniform to preserve the normal beats distribution at 66.67%.

(b) Support Vector Machine

We use a SVM based classifier, to generate support vectors, which are then used for classification. We chose SVM due to its obvious advantages, such as - its effectiveness in high dimensional spaces, the versatility with various kernels, etc.

We initially used **58** of the described 68 features. A classifier was built [4], and the testing revealed 92.2% accuracy. All features were normalized, and standardized (Zero mean, Unit Variance). Ten more features were added to reach **68** and the accuracy improved to **94.6%**. Here, a polynomial 3rd degree (optimum currently) kernel was employed.

6 Feature Selection

To improve performance, we calculate feature importance using an ensemble learning method “Random Forest” [5]. Here the algorithm constructs levels of decision trees, and the output is the mode of the classes output by individual trees. Each tree fits the data to a randomized subspace. By perturbing and conditioning the node independently,

it accumulates a bunch of estimates. The motive is that, given a large number of nodes, the errors tend to even out.

Here, we obtain the feature importance vector, that quantifies the weight of each feature. From the performance plot Figure[5], we choose the top **49** features. Even though, there are lower feature size with comparable error, being generous with the choice here may avoid under fitting in the subsequent stages. Note that the training error falls and saturates after some point, so 49 is not a bad choice as the data has not been over fit yet.

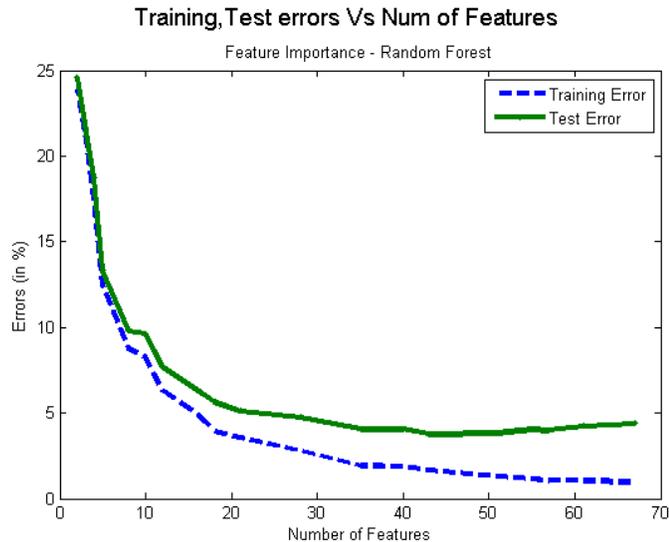


Figure 5: Random Forest Selection

Now, the performance is **96.2%** on test, and **98.5%** on training sets.

We also explored two other classes of features selection. Filter feature selection attempts to evaluate the “helpfulness” of a feature independently of the algorithm which will ultimately be used to generate a hypothesis. Then, we used Mutual Information as our scoring metric.

7 Training size

Now, the classifier (SVM 3rd order) is now iterated to find the optimal training size. Figure[6] shows that the test error doesn’t improve after ~ 9000 samples. Note the training error falls below (over-fit) 8000, and over 10000, so the size [8000,10000] seems to be perfect size for the classifier. So we retain our training size of **10000**.

8 Dimensionality

We now use Principal component analysis to find a reduced subspace. We compute the top few eigen values, and corresponding eigen vectors to linearly combine the vectors to create new feature matrix.

To find the optimal number of dimensions, we iterate the classifier with 49 features over a PCA routine, and Figure

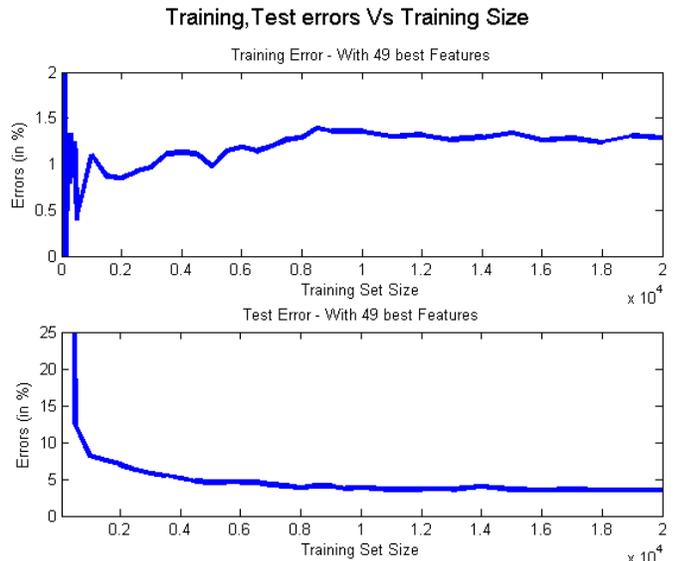


Figure 6: Training Size Optimization

[7] gives useful insights on the trend. Note the training error fall(over-fit) and rise(bias) again. The test error minimum occurs at **38** dimensions, and the training error is not too high nor low, and the test performance is **96.84 %**

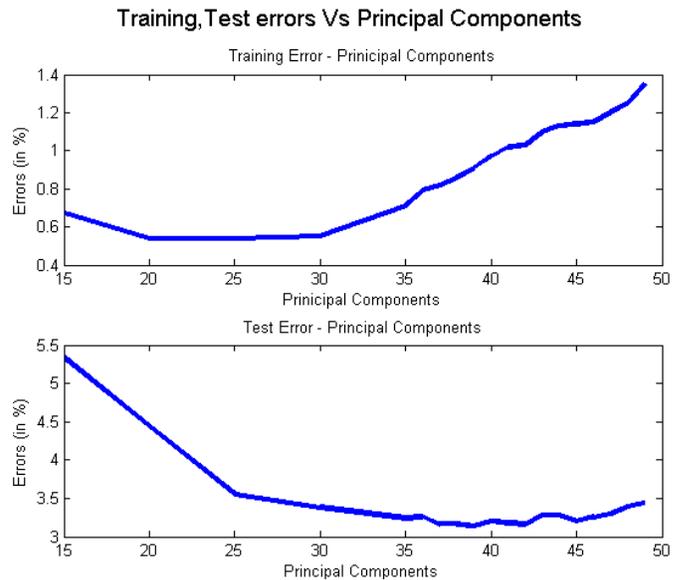


Figure 7: Dimensionality Reduction

9 Kernel selection

Till now, we have been using a third order polynomial kernel. On iteration, we find that 4th degree polynomial gives a maximum performance of **97.2%**(Figure [8]).Note the training error is zero over degree 5, which is clear over-fitting, and note high bias at low degrees. The possibility of this optimal point to be local is really high, given the arbitrariness in the design flow. However, the improvements in a different classification loop will be trivial for the given set of features.

Training, Test errors Vs Polynomial degree - Intercept-1.0

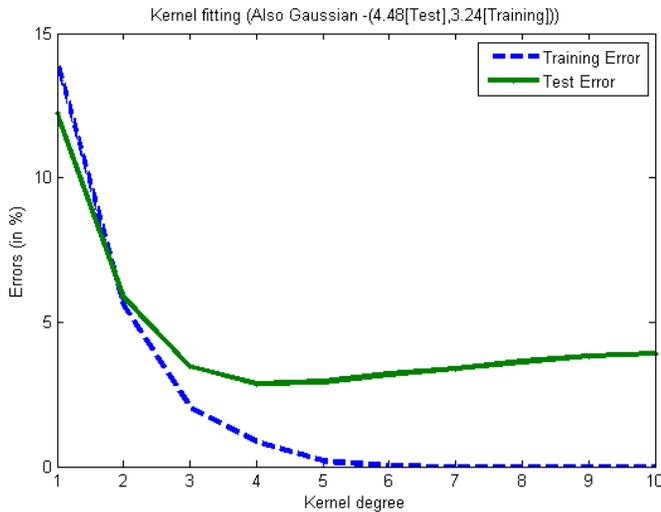


Figure 8: A single EKG pulse example [1]

The kernel intercept is chosen to be 1.0 after optimizing that over a set of values in $[0.2, 2.5]$.

10 Results

Table [1] provides the performance at every stage of the learning process.

1. The top four features are 1st wavelet coefficient, Pre-Peak (P-wave) energy, 2nd wavelet coefficient, and amplitude ratio of QRS/P.
2. the final confusion matrix elements are 65.66% True Positive, 31.18% True negative, 1.9% False positive, and 1.26% False negative, for a distribution of 66.67% normal beats.
3. Few {performance, # of top features} tuples are {78.1%,2}, {82.3%,4}, {86.5%,6}, and {91%,9}.

Step	Features #	Training Performance(%)	Test Performance(%)	Optimum
Initial	58	94	92.2	–
Add Features	68	95.6	94.6	–
Random forest	49	98.5	96.2	–
Train Size	49	98.5	96.2	10000
PCA	38	98.7	96.84	–
Kernel Select	38	99.14	97.2	Poly 4

Table 1: Performance at every stage

11 Conclusion

The above procedure was tested on multiple test data, each with 5000 - 10000 samples, and the results closely match in all training sets. to the order of $\pm 0.4\%$.

In practice, the testing beats are not sampled from various records, but the learning algorithm must be run on a single record, and predictions must be made. That case,

it is important to have a set of global features compliment the decision, as local beat behavior may not satisfactorily imply the global health. Clearly, there is immense scope to improve the feature extractions, and selections.’

References

- [1] M. K. Slate. (2013) EKG interpretation. coursematerial-187.pdf. [Online]. Available: <http://www.rn.org/courses/>
- [2] J. M. Prutkin. (2013) ECG tutorial: Basic principles of ECG analysis. [Online]. Available: <http://www.uptodate.com/contents/ecg-tutorial-basic-principles-of-ecg-analysis>
- [3] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. [Online]. Available: <http://www.physionet.org/physiobank/database/mitdb/>
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [5] L. Breiman, “Random forests,” in *Machine Learning*, 2001, pp. 5–32.