Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier

10 authors, including:

Chetanya Puri
KU Leuven
21 PUBLICATIONS **153** CITATIONS

SEE PROFILE

Ayan Mukherjee
Indian Institute of Technology Kharagpur
13 PUBLICATIONS **83** CITATIONS

SEE PROFILE

Rohan Banerjee
Tata Consultancy Services Limited
41 PUBLICATIONS **282** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Connected Digital Health View project

Project   Domain Specific Engineering Platform View project

Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier

Shreyasi Datta, Chetanya Puri, Ayan Mukherjee, Rohan Banerjee, Anirban Dutta Choudhury
Rituraj Singh, Arijit Ukil, Soma Bandyopadhyay, Arpan Pal, Dr Sundeep Khandelwal

Tata Consultancy Services

Abstract

In this paper, we present a methodology for classifying normal, atrial fibrillation (AF), non-AF related other abnormal heart rhythms and noisy recordings by analysing single lead ECG signal of short duration. In a two layer binary cascaded approach proposed in our methodology, an unlabelled recording is initially classified into one of the two intermediate classes ('normal+others' and 'AF+noisy') at the first layer before actual classification at the second layer. The Physionet Challenge 2017 dataset containing more than 8500 ECG recordings are used for creation of training models and interval validation. The proposed methodology yields an average F1-score of 0.91, 0.79 and 0.77 respectively in classifying normal, AF and other rhythms on the training dataset using 5-fold cross validation. Results also show that, the said methodology, when applied on a hidden test set maintained by the challenge organisers yields F1-score values of 0.92, 0.86 and 0.74 in classifying the same.

1. Introduction

Atrial Fibrillation (AF) is a common type of heart disease that leads to stroke, heart failure or other complications. Millions of people get affected by AF every year and the prevalence of the disease is likely to increase. Non-invasive detection of AF is a popular area of research for quite a long time. Irregularities in heart beat is considered to be the most common symptom of AF and can be traced in an ECG. However, being an episodic event an accurate detection of AF is not always trivial. Conventional AF detectors ([1], [2], [3]) are mostly of atrial activity analysis-based or ventricular response analysis-based methods. The absence of P waves or the presence of f waves in the TQ interval are searched in atrial activity analysis-based AF detectors. On the other hand, time, frequency and morphological features are extracted from RR intervals to identity the heart beat irregularity in ventricular response analysis-based methods.

However, the prior art methods have certain limitations regarding real time deployments. 1) Most of them are validated on clinically accepted 12 lead ECG signals, recorded for a relatively longer duration. 2) Algorithms are mostly applied on carefully selected clean data. However, in practical scenario, ECG signals are often noisy in nature. 3) Size of the test dataset are often not adequate for making a conclusion. 4) Most prior arts perform binary classification between AF and normal recordings only. However, there are many non-AF abnormal rhythms (like tachycardia, bradycardia, arrhythmia etc) which exhibits heart beat pattern similar to AF. Considering them in the in the dataset makes the classification task more challenging.

In this paper we propose a robust algorithm for classifying normal, AF, other abnormal rhythms and noisy ECG recordings. The diverse ECG dataset, provided in Physionet challenge 2017 [4] is used for internal performance evaluation and creating the training models. Information regarding individual recordings are not available regarding the other rhythms in the dataset as all non AF related abnormal rhythms are clubbed together to form a single class. Our contributions in this paper are, 1) introducing a multi-layer cascaded binary classifier instead of a single multi-class classifier, 2) front end noise removal and 3) feature selection at each layer of classification form a pool of more than 150 features.

2. Methodology

A multi-layer cascaded binary classification approach is incorporated in our methodology. Prior arts [5], [6] show that for an unbalanced multi-label classification problem, a series of efficiently chosen cascaded binary classifiers often outperforms a single multi-class classifier. As shown in Figure 1, three classification jobs are performed at two layers connected in cascade. An unknown recording is temporary marked as *'normal+ other rhythms'* or *'AF + noisy'* at layer 1. The four label classification is performed at layer 2. A more detailed discussion of different components of the methodology is provided subsequently.
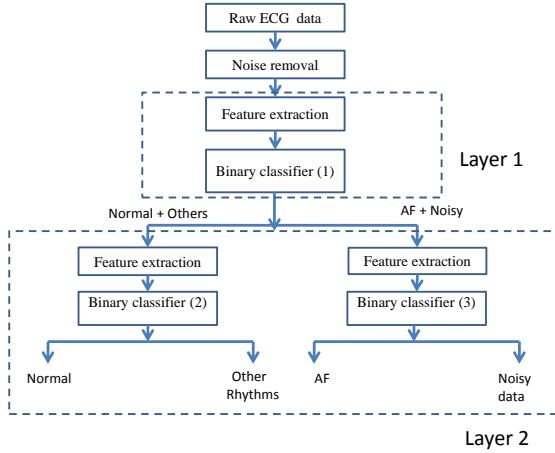
Figure 1. Architecture of proposed solution.

2.1. Noise removal

ECG signals, recorded using single lead, non-medical grade equipments are inherently noisy in nature. Body movement of the user during recording, voltage fluctuation of the sensor device or improper contact between subject body and the sensor electrodes can heavily corrupt the signals. A low frequency component is also present due breathing. Locating and discarding the noisy portions inside a signal is considered a prerequisite before feature extraction and classification.

In this paper we implement a simple spectrogram based approach for identifying the noisy parts inside an ECG recording.
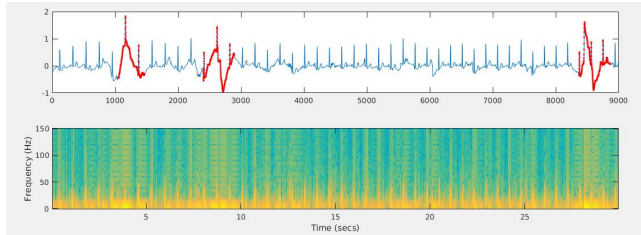

Figure 2. Spectrogram based noise removal.

Spectrogram of the entire signal is computed for analysis. Important cardiac information, are typically stored within 20 Hz in ECG waveform. Thus, following the spectrogram plot across the time axis, if we can locate the regions between successive RR intervals, having high spectral power above 50 Hz (as shown in Figure 2), that can be safely marked as noisy and are discarded before further processing. Further, the clean signal is passed through a high pass filter (cut-off frequency 0.5 Hz) to remove the baseline movement of the signal. If more than 80% of a certain waveform is discarded in this process, it is marked

as noisy, bypassing the two layer classification job.

2.2. Feature extraction

Identifying the differentiating feature set is considered the most important task of any classification problem. An accurate detection P, Q, R, S, T points are the the primary requirement for an accurate feature extraction. A modified version of Pan-Tompkins algorithm, combining with several open source algorithms is implemented to identify QRS complex and R peaks. In the subsequent stage P,Q,S and T points are located, taking R peaks as reference points (as shown in Figure 3). A wide list of more than 150 features are considered in our paper. The features can be broadly classified into following categories.
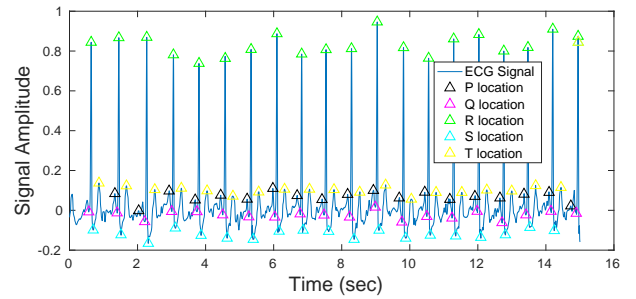

Figure 3. PQRST detection on single lead ECG recording.

2.2.1. Morphological ECG features

These features are derived from the PQRST points detected in an ECG waveform. These features include the median, range and variance of the corrected QT interval (QTc), QR and QRS widths, slopes of QR, RS and ST intervals, depth of the Q and S points with respect to R, amplitude difference of the TR wave; ratio of the number of P waves to the number of R waves and distance of the ST segment crossing from the S point among others. These features are widely used by clinicians for identifying cardiac abnormalities. It is to be noted that some of these features are indicative of AF while others are for detecting the other abnormal rhythms. For example, in AF, P waves are often absent [7], hence the ratio of number of P waves to the number of R waves detected is supposed to be very small. Hence, this particular feature was selected by our automated feature selection methodology for the Layer 1 classifier and the classifier 3 of layer 2, however, was not selected for the classifier 2 of layer 2.

2.2.2. Prior art AF features

Several features are widely available in prior arts, to identify AF events. Irregular RR intervals is a very common symptom present in AF patients. There are several metrics to identify the same. Sarkar *et al.* [2] pro-

posed several measurements like AFEvidence, Original Count, Irregularity Evidence, Pace Count, Density Evidence, Anisotropy Evidence, AFEvidence from Lorentz plot of RR intervals. Certain features [1] are derived from the the inter beat intervals using Poincare plots also reported commending accuracy in identifying AF. Other features include approximate and sample entropy based features for AF detection [3], [8] as well as coefficient of variation of RR and delta RR intervals.

2.2.3.   HRV features

Several HRV related features are also incorporated in our analysis. features like pNNx (number of NN intervals above x, normalized by duration of recording, where x lies between 20 and 500 ms), SDNN, SDSD and normalized RMSSD are widely used in literature in identifying cardiovascular diseases [9]. Normalized spectral power of the RR interval time series within the frequency region of 0-0.04 Hz, 0.04-0.15 Hz and 0.15-0.5 Hz are also used.

2.2.4.   Frequency features

We have explored certain frequency domain features widely used in many biomedical and other applications. Raw time signal is broken into small windows of 2 seconds duration having 50% overlapping using hamming window. Frequency analysis is performed by computing Short Time Fourier Transform (STFT) of each window. Extracted features are, mean spectral centroid, spectral roll-off, spectral flux along with normalized spectral power between 0-10 Hz and 10-20 Hz across all windows in a measurement.

2.2.5.   Statistical features

Statistical features including mean, median, variance, range, kurtosis and skewness of RR intervals and the probability density estimate (PDE) of the RR intervals and the delta RR intervals. Additionally, the number of peaks on the PDE of the RR and delta RR intervals along with the variation of energy in between the RR peaks are also used as features. The Shannon, Tsallis and Renyi entropy, Linear Predictive Coefficients (LPC) of the raw time series data are also incorporated in our feature list.

2.2.6.   Features for other abnormalities

Here we try to extract some of the major features to detect arrhythmia or other non-AF related abnormalities present in an ECG signal. Various features used are extracted using sliding window with six peaks per window and its average RR interval, maximum of first difference of number of samples in the window with magnitude exceeding 0.1mV, the normalized power spectrum density (nPSD)

of the window, etc. The heart rate was estimated using an adaptive frequency tracking algorithm (AMM) to derive the features as mean of RR interval, decrease of HR, max SPI index, average HR, abnormal HR etc.

2.2.7.   Features for detecting noisy recordings

There are several works that detect noise/motion-artifacts in the different portions of the signals. However, there are times when the complete signal is useless for cardiac abnormality derivation. We use domain dependent time and frequency features [10] along with certain statistical features that exploit the rise and fall [11] in the morphology of the ECG signal for improved noise detection. These features discriminate well between the regularities of the clean ECG signal versus the randomness in a noisy waveform.

2.3.   Feature Selection

Feature selection often improves classification accuracy by removing the noisy (irrelevant) features and also reduces the computation time by removing redundant features. A separate feature set is used in each of the three classifiers of our methodology, efficiently chosen using statistical feature selection tools like Maximal Information Coefficient (MIC) and minimum Redundancy Maximum Relevance (mRMR) and are finalized during training phase. Features typically recommended by medical practitioners for a certain level of classification are also incorporated in the selected feature list.

2.4.   Classification

AdaBoost (adaptive boosting) is used in all levels of classification. Adaboost is an ensemble learning approach, where many weak learners are iteratively added in each round of training and a weighting vector is adjusted to reduce the misclassification rate. This algorithm is less prone to overfitting, more sensitive to noisy data and outliers and can handle the class label imbalance in the training set. For each of the three classifiers, two parameters of the ensemble classifiers, namely, number of learning cycles and learning rate are optimized using Bayesian optimization function.

3.   Experimental results

All our experiments are performed on the Physionet 2017 challenge dataset [4]. This dataset comprises a total of 8528 recordings with mean duration of 32.5 seconds. All of them are single lead ECG signals, recorded using AliveCor device and sampled at 300 Hz. The distribution

of normal, AF, other rhythms and noisy data is largely imbalanced in the dataset. F1-score for normal (F_{norm}), AF (F_{af}) and other rhythms (F_{oth}) are computed for performance analysis and the final accuracy is reported in terms of $F_T = (F_{norm} + F_{af} + F_{oth})/3$.

Our proposed methodology is initially applied on the entire training dataset using 5-fold cross validation and the performance of the same is shown in Table 1.

Table 1. Performance of proposed methodology on training dataset using 5-fold cross validation

	F_{norm}	F_{af}	F_{oth}	F_T
Average	0.9095	0.7978	0.7719	0.8264
Std.	0.0022	0.0062	0.0040	0.0026

It can be observed that the performance of our methodology is stable enough across the folds in a 5-fold cross validation, owing to the small standard deviation values across different test scenarios. It can also be observed that the methodology yields a very high F1-score (average value 0.9) in classifying normal recordings, where as the F1-score of detecting other rhythms is close to 0.77. A detailed inspection reveals that a good amount of other rhythms are wrongly classified as normal. The unavailability of actual disease information corresponding to each recording in this particular class makes the classification task more challenging to identify the proper features, thus reducing the classification accuracy.

Our final training model is created on the entire training dataset and evaluated on a subset of the hidden test dataset at a server maintained by Physionet challenge organizers.

Table 2. Performance of proposed methodology on a subset of hidden test dataset

Data set	F_{norm}	F_{af}	F_{oth}	F_T
Complete training data	0.99	0.94	0.98	0.96
Complete test data	0.92	0.82	0.75	0.83

Table 2, shows the performance of our methodology on the complete training data made available during the challenge and also the test data, kept hidden by the conference committee for evaluating different submitted algorithms [4] for ranking. Our methodology ended up securing the top position in the challenge with overall F1-score of 0.83 [4].

4. Conclusion

In this paper we have proposed a methodology for classifying short single lead ECG recordings using a series of multi-layer cascaded binary classifiers. The methodology has been successfully validated on Physionet 2017 dataset. Our methodology is available as a open source software under the terms of the GNU General Public License as published by the Free Software Foundation. Our future work plan includes identifying some more significant features for improving the accuracy of other rhythms. We are also planning to apply deep learning tools on the training set for an improved classification accuracy.

References

[1] Park J, Lee S, Jeon M. Atrial fibrillation detection by heart rate variability in poincare plot. Biomedical engineering online 2009;8(1):38.

[2] Sarkar S, Ritscher D, Mehra R. A detector for a chronic implantable atrial tachyarrhythmia monitor. IEEE Transactions on Biomedical Engineering 2008;55(3):1219–1224.

[3] Lake D, Moorman JRM. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. American Journal of Physiology Heart and Circulatory Physiology 2011;300(1).

[4] Clifford G, et al. Classification from a short single lead ecg recording: the physionet computing in cardiology challenge. In Computing in Cardiology Conference (CinC), 2017. IEEE; .

[5] Doyle S, et al. Cascaded multi-class pairwise classifier (cascampa) for normal, cancerous, and cancer confounder classes in prostate histology. In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. March 2011; 715–718.

[6] Kopinski o. A pragmatic approach to multi-class classification. In Neural Networks (IJCNN), 2015 International Joint Conference on. IEEE, 2015; 1–8.

[7] Prerfellner H, et al. P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors. Heart Rhythm 2014;11(9):1575 – 1583.

[8] Alcaraz R, et al. Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. Computer Methods and Programs in Biomedicine 2010; 99(1):124 – 132.

[9] Banerjee R, et al. Identifying coronary artery disease from photoplethysmogram. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM; 1084–1088.

[10] Puri C, et al. Classification of normal and abnormal heart sound recordings through robust feature selection. In Computing in Cardiology Conference (CinC), 2016; .

[11] Bandyopadhyay, et al. An unsupervised learning for robust cardiac feature derivation from ppg signals. In EMBC. IEEE, 2016; .

Address for correspondence:

Shreyasi Datta

Embedded Systems and Robotics, Tata Consultancy Services
Building 1B, Ecospace Plot - IIF/12, New Town, Rajarhat
Kolkata - 700160, India
shreyasi.datta@tcs.com