

Classification and Diagnosis of Cardiac Arrhythmia using an ECG-based Ensemble Approach

¹Ank Agarwal, ²Akshar Agarwal

¹*Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America;*

²*Neurology, Yale University School of Medicine, New Haven, Connecticut, United States of America;*
aagarw26@jhu.edu; akshar.agarwal@yale.edu

ABSTRACT

Cardiovascular Disease (CVD) remains the leading cause of death, worldwide and in the United States. Approximately 30% of global deaths can be attributed to one form of CVD, including conditions such as heart disease, stroke, heart attack, and arrhythmia. In diagnosing CVD, electrocardiograms (ECG) are commonly used to measure and record the electrical activity of the heart. Their non-invasive, informative, and relatively simple nature allows for rapid deployment. However, because analysis of ECGs depends solely on a physician, ECG analysis becomes subjective, adding a potential layer of error to patient healthcare. Studies indicate that physicians often misread ECGs and disagree with each other's interpretations. In order to develop an accurate and objective method for ECG analysis, this study evaluates various ensemble algorithms to design and create a supervised classification model. Several ensemble models were evaluated to derive one which correctly classifies CVD with sufficiently high accuracy. A boosted decision tree ensemble created to evaluate cardiac condition performs best, with an overall accuracy of 84.6% and an AUC of 0.828.

Keywords: computer-aided diagnosis, arrhythmia, AI-based clinical decision making

1 Introduction

Cardiovascular disease (CVD) is the number one cause of death both globally and in the United States, accounting for 17.3 million deaths per year (30% of all deaths). In fact, it is estimated that the number will rise to 23.6 million deaths per year by 2030 [1].

In order to diagnose CVD, physicians utilize electrocardiogram (ECG) recordings. ECGs provide measurements, indicating the electrical activity of the heart. Despite CVD being the most prevalent cause of death worldwide, it is diagnosed using physician-based analysis of ECG data, which can potentially add subjective interpretation of patient data. In the United States, most ECGs are read by non-cardiologists with nominal training in ECG readings [2]. Furthermore, non-cardiologists often disagree on ECG analyses [3]. In the emergency room, errors in ECG analysis can delay appropriate treatment of patients at hand, as well as result in delayed treatment of patients waiting in the queue [4]. Thus, there is a clear and present need for an objective, accurate, and rapid method of ECG analysis.

Several attempts have been made in the past to produce a capable classification model for arrhythmia. Some of these studies focused on individual classifiers, such as relevance vector machines (RVM) [5]. Other approaches have attempted to utilize artificial neural networks (ANN), Markov chains, and support vector machines (SVM). Machine learning models have been applied to arrhythmia classification in the past, however the novel approach described herein utilizes an ensemble model. This ensemble approach offers a more robust model with reduced error. Therefore, multiple ensemble approaches were evaluated to accurately predict heart condition from electrocardiogram data.

2 Methods

2.1 Dataset

The dataset used herein was obtained from the University of California (UCI), Irvine Machine Learning Repository, contributed by Guvenir et al. [6,7]. The dataset contains information from 452 patients, consisting of 279 attributes. The patients' conditions are labeled as one of 16 conditions, 15 irregular and one healthy. The features primarily consist of ECG data: heart rate, PQRST wave signals, channel information, et cetera. Other features are general subject information: age, sex, height, weight, et cetera. Table 1 below shows the categories/conditions and their respective sizes from the initial study. The data was largely intact, but several datum points were missing. The 14th feature, which consisted of J point data, was missing for the overwhelming majority of patients; of the 452 patients, 376 patients had missing values for feature 14. Data from feature 14 was therefore excluded from all patients. Additionally, 32 patients had missing values in a variety of features. Therefore, instead of interpolating those data, the 32 subjects were removed altogether. Instances which were labelled as 16, miscellaneous heart conditions, were also removed. The dataset used for the machine-learning model described herein consists of 402 patients with 278 features/attributes, belonging to 11 classes. Two-class performance (healthy or ill) was also assessed for each model.

Table 1: Instances of arrhythmia and their respective counts in the original dataset

Condition Code	Condition	Number of instances
1	Healthy	245
2	Ischemic changes (Coronary Artery Disease)	44
3	Old Anterior Myocardial Infarction	15
4	Old Inferior Myocardial Infarction	15
5	Sinus tachycardia	13
6	Sinus bradycardia	25
7	Ventricular Premature Contraction (PVC)	3
8	Supraventricular Premature Contraction	2
9	Left bundle branch block	9
10	Right bundle branch block	50

11	1st degree AtrioVentricular (AV) block	0
12	2nd degree AV block	0
13	3rd degree AV block	0
14	Left ventricular hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

2.2 Machine Learning Model

Decision trees are a standard supervised learning method and are often used for both classification and regression. Classification and Regression Trees (CART) are relatively simple and easy to implement, but come with many tradeoffs. CART models introduce high variance and are quite unstable individually. To reduce the variance that a single unstable CART model might incur, various ensembles are proposed here. Several forms of boosting and aggregation are evaluated.

Boosting is an additional method of combining individual classifiers through the specification of a cost function and majority voting. It is also well suited for decision trees. RUSBoost, a popular boosting algorithm, rectifies class imbalances by drawing a subset of samples from each class and then follows traditional reweighting procedures for creating an ensemble [8-11].

2.3 Implementation

Using the MATLAB Statistics and Machine Learning Toolbox, the ensemble approach was used here as it is favorable to individual CART models due to reductions in variance and overall increases in accuracy. In order to evaluate various ensembles and perform model selection, Bayesian optimization was used as described by Snoek, J et al. [12]. Bayesian optimization explores various hyperparameters in search of minimizing the objective function. This method of model and hyperparameter selection is favorable because grid searches are computationally intensive, and although random searches have been demonstrated to reduce error when certain criteria are met, Bayesian optimization produces non-trivial values in short time. A large quantity of tunable parameters exist, including, but not limited to: methods of boosting/bagging, maximum depth, minimum number of splits, split criterion, et cetera. As the number of tunable parameters increases, the time required for a grid search exponentially increases, leaving Bayesian optimization to be the preferred method. Bayesian optimization entails training a Gaussian process and computing the expected improvement. A model is constructed through exploration of certain values and predictions are made regarding which values will return the greatest expected improvement.

3 Results

Various ensembles were evaluated, as shown in Table 2.

Table 2: Performance Statistics for all evaluated ensemble algorithms

11-Class						
Data	Overall Accuracy	C-Statistic (AUC)	Sensitivity	Specificity	Predictive Value	
					Positive	Negative
Random Forest	0.8202	0.8449	0.7576	0.9241	0.8741	0.8456
RUS Boosted	0.2363	0.5524	0.3714	0.7706	0.3421	0.7925
Subspace Discriminant	0.5970	0.8059	0.8333	0.9033	0.4902	0.9798
Subspace KNN	0.5597	0.6563	0.8333	0.7771	0.1765	0.9879
2-Class						
Data	Overall Accuracy	C-Statistic (AUC)	Sensitivity	Specificity	Predictive Value	
					Positive	Negative
Random Forest	0.8458	0.8278	0.7273	0.9283	0.8759	0.8302
RUS Boosted	0.8259	0.8841	0.7975	0.9707	0.9489	0.8755
Subspace Discriminant	0.7736	0.7454	0.7833	0.8475	0.6861	0.9019
Subspace KNN	0.7139	0.6718	0.7553	0.7857	0.5182	0.9132

Bayesian optimization was iteratively performed 30 times per classifier and the configurations with the highest accuracies were evaluated. The highest 11-Class accuracy was achieved with a random forest, an extension of bootstrap aggregation [13]. Random forests have been applied to a variety of biological problems [14-16]. This method involves utilizing several hundred individual trees to reach a classification. A random sample of features/predictors is utilized for each individual tree. Here, the random forest with 128 weak learners (trees) operates with 82.02% accuracy and has an AUC of 0.8449. The best 2-Class classifier was determined to be a boosted trees ensemble utilizing RUSBoost, an algorithm which alleviates class imbalances to obtain higher accuracy [17]. The boosted trees had an overall accuracy of 82.59% accuracy and an AUC of .8841. Figure 1 shows the Receiver Operating Characteristic (ROC) curve for both the optimal 11-Class and 2-Class models.

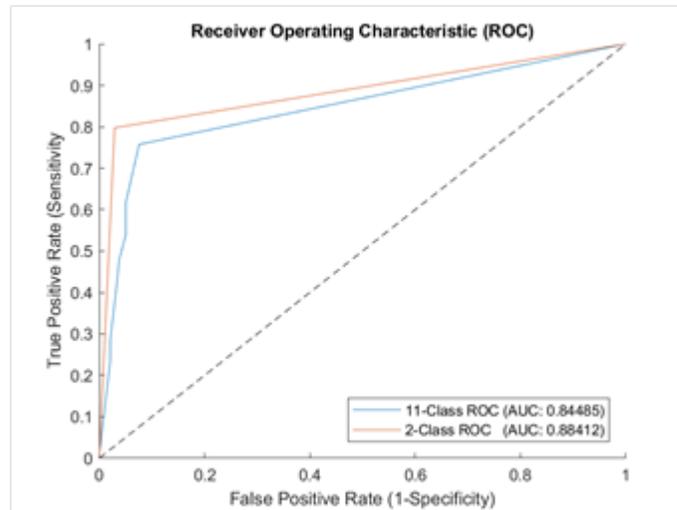


Figure 1. Receiver Operating Characteristic (ROC) curve for both the 2-class and 11-class models, detailing the relative performances of each model.

4 Discussion

The optimal model generated here, trees boosted with RUSBoost, has a specificity of 97.07%, sensitivity of 79.75%, and PPV of 94.89% when classifying heart irregularities. This shows vast improvement over both other computer models and physician-based interpretations of arrhythmia-related ECGs.

In comparison to previous machine learning projects and physician-gauged studies aimed at analyzing ECGs for cardiovascular conditions such as arrhythmia and heart failure, the algorithm presented herein had greater sensitivity, specificity, and positive predictive value (PPV). Sensitivity quantifies the ability of the proposed model to correctly detect heart irregularities amongst patients with heart irregularities. Similarly, specificity quantifies the ability of the model to detect when patients are healthy.

Compared to algorithms which utilized voting feature intervals, naive Bayes, k-nearest neighbors (KNN), support vector machines (SVM), and logistic regression, the model proposed here offers increased accuracy [6,18-20].

Additionally, both the random forest and the boosted trees offer the advantage of reduced variance given that it combines predictions from many independent weak learners. The previous studies also often utilized various cross validation methods (different training/testing splits or folds).

Furthermore, the proposed method also performed better than machine learning algorithms developed for analyzing other heart conditions, such as heart failure, with improved PPV [21]. The area under the ROC curve represents how well a classifier can distinguish between two classes (healthy or diseased).

Compared to healthcare professionals, the 2-class algorithm developed in this study had better specificity and accuracy. For instance, it outperformed cardiologists, general physicians, and general practice nurses in terms of specificity when diagnosing cardiac well-being [22]. Moreover, our false positive rate for ECG-based decision-making was 2.93%, compared to numbers as high as 41% for emergency medicine professionals when decisions needed to be made quickly [4]. Furthermore, physicians often disagree on ECG interpretations, with kappa values among cardiologists of 0.69 and among general physicians of 0.52 [3]. Kappa values indicate the rate of agreement amongst different

sources for categorical variables, and in this case, how often different doctors agreed on ECG-based diagnoses. Inaccuracies and disagreements not only delay proper treatment of patients at hand, but result in excess hospital admissions, thus delaying treatment for others. Therefore, usage of a computer algorithm for ECG-analysis would solve not only issues of accuracy, but also speed of diagnosis.

5 Conclusion

The best performing ensemble method proposed here functions with approximately 82.59% accuracy. Although this ensemble didn't have the highest accuracy, its sensitivity, a more transferable metric in clinical settings, was particularly high. The performance is a vast improvement over other classifiers and healthcare professionals. The highest performing ensembles tested here were the random forest and boosted trees. The ensembles, favorable to individual classifiers due to their impactful reduction on variance and improved accuracy, perform with high accuracy, sensitivity, and specificity, allowing for a proper classification of cardiological conditions in real-time.

6 Future Directions

More changes made in the preprocessing phase can have large impacts on the end result. Rather than removing certain patients' data and features, it may prove useful to interpolate certain datum points, collect more refined data, and investigate feature selection. Furthermore, the model developed in this experiment could be improved by introduction of further data, and a validation phase against data from outside the original data set.

REFERENCES

- [1] Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics— 2015 update: a report from the American Heart Association. *Circulation*. 2014; 131(4):e29-322.
- [2] Drew, B., Dracup, K., Childers, R., Criley JM, Fung G, Marcus F, et al. Finding ECG Readers in Clinical Practice. *J Am Coll Cardiol*. 2014;64(5):528.
- [3] Magee C, Kazman J, Haigney M, Oriscello R, DeZee KJ, Deuster P, et al. Reliability and Validity of Clinician ECG Interpretation for Athletes. *Ann Noninvasive Electrocardiol*. 2014;19(4):319-329.
- [4] Jayes RL, Larsen GC, Beshansky JR, D'Agostino RB, Selker HP. Physician electrocardiogram reading in the emergency department—Accuracy and effect on triage decisions. *J Gen Intern Med*. 1992;7(4):387-392.
- [5] Gayathri S, Suchetha M, Latha V. ECG Arrhythmia Detection and Classification Using Relevance Vector Machine. *Procedia Eng*. 2012;38:1333-1339.
- [6] Güvenir HA, Kaçar M, Demiroz G, Cekin A. A supervised machine learning algorithm for arrhythmia analysis. *Computers in Cardiology*. 1997;24:433 - 436. 10.1109/CIC.1997.647926.
- [7] Dua D and Karra Taniskidou E. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. 2017.
- [8] Mathworks. Statistics and Machine Learning Toolbox: User's Guide (r2018a). Retrieved August 01, 2018 from https://www.mathworks.com/help/pdf_doc/stats/stats.pdf

- [9] Schapire RE, Freund Y, Bartlett P, Lee WS. Boosting the margin: a new explanation for the effectiveness of voting methods. *The Annals of Statistics*. 1998;26(5):1651-1686.
- [10] Friedman J, Tibshirani R, Hastie T. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*. 2000;28(2):337-407.
- [11] Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. Lecture presented at the: 1999; Institute of Mathematical Statistics.
- [12] Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. arXiv:1206.2944 [stat.ML]
- [13] Breiman, L. *Machine Learning*. 2001; 45: 5.
- [14] Gray K, Aljabar P, Heckemann R, Hammers A, Rueckert D. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage*. 2013;65:167-175.
- [15] Svetnik V, Liaw A, Tong C, Culberson J, Sheridan R, Feuston B. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J Chem Inf Comput Sci*. 2003;43(6):1947-1958.
- [16] Chen T, Cao Y, Zhang Y, Liu J, Bao Y, Wang C, et al. Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection. *Evidence-Based Complementary and Alternative Medicine*. 2013;2013:1-11.
- [17] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. 2010;40(1):185-197.
- [18] Zuo WM, Lu WG, Wang KQ, Zhang H. Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier. *Computers in Cardiology*. 2008; 35:253-256.
- [19] Uyar A, Gurgen F. Arrhythmia Classification Using Serial Fusion of Support Vector Machines and Logistic Regression. 2007 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. 2007:560-565
- [20] Soman T, Bobbie PO. Classification of Arrhythmia Using Machine Learning Techniques. *WSEAS Transactions on Computers*. 2004;4(6).
- [21] Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, et al. Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. *JAMA Cardiol*. 2016;1(9):1014.
- [22] Shiyovich A, Wolak A, Yacobovich L, Grosbard A, Katz A. Accuracy of Diagnosing Atrial Flutter and Atrial Fibrillation From a Surface Electrocardiogram by Hospital Physicians: Analysis of Data From Internal Medicine Departments. *The American Journal of the Medical Sciences*. 2010;340(4):271-275.