



## Supervised Machine Learning Algorithms for Arrhythmia Classification and Diagnosis

---

Avishek Choudhury and Nikhli Shetty

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 22, 2020

# Supervised Machine Learning Algorithms for Arrhythmia Classification and Diagnosis

Avishek Choudhury<sup>1a</sup> and Nikhil Shetty<sup>1</sup>

<sup>1</sup>Stevens Institute of Technology, Hoboken, NJ

<sup>1a</sup>achoudh7@stevens.edu

## Abstract

Preventing arrhythmic risks in patients with cardiac ailment has been a global concern. No reliable method for diagnosing Arrhythmia exist. Machine learning techniques can be used to predict the advent of arrhythmic risk and recommend appropriate measures to ensure patient safety. In this paper, we implement seven machine learning algorithms to predict Arrhythmia. Our model promises to produce better results than the existing VF15 algorithm and cardiologists.

Additionally, our study also advocates that machine learning in arrhythmia identification can reduce diagnostic expenses by minimizing type-I and type-II errors. Our study implements a random forest, gradient boosted trees, artificial neural networks, support vector machine, XG-boost, logistic regression, and ensemble method. All the models in this study have been evaluated using classification accuracy, precision, recall, F1 score, gain, lift, ROC, and confusion matrix. A comparative analysis of all the models highlights the strengths and weaknesses of individual algorithms. The ensemble method yields the highest accuracy (0.84), gain (0.31), F1 score (0.81), and ROC (0.87). It produces moderately high false-negative (14%), whereas, support vector machine with an accuracy of 0.71, generates the highest recall value of 0.91 (sensitivity) and a minimum false-negative (11%). The best performing model in this study outperforms the accuracy exiting VF15(62%) technique by more than 20% margin and sensitivity of cardiologist (0.78) by 13%.

## Keywords

Arrhythmia, machine learning, artificial intelligence, artificial neural network, support vector machine.

## 1 Introduction

Arrhythmia is a cardiac disorder indicating an irregular heartbeat rhythm. In this ailment, the normal beating in the heart's upper atrium gets distorted [1], affecting the flow of blood to the lower chambers of the heart. Arrhythmia is a leading cause of death every year. The number of patients susceptible to arrhythmic risk with a Sudden Cardiac Death (SCD) is increasing [2, 3]. According to the Centers for Disease Control and Prevention (CDC), an estimated amount of \$2.7-\$6.1 million has been spent on treating people with Arrhythmia in the USA. Moreover, Arrhythmia has been responsible for death of about 130,000 USA residents and 750,000 hospitalizations every year. Sudden cardiac death (SCD) cause due to Arrhythmia represent a global public health concern, accounting for 15–20 % of all deaths [4].

Most traditional diagnosis methods are flawed and often result in delayed treatment or misdiagnosis. Due to the complexity of arrhythmia [5], no single test reliably predicts arrhythmic risk in patients with heart failure [6]. A combination of tests improves risk stratification and helps identify patients at higher risk of SCD [6]. However, multiple tests increase diagnostic expenses. Electrocardiogram (ECG) is one of the most common procedures used to diagnose this disease. However, this method has many drawbacks [7]. The errors in the diagnosis of Arrhythmia and other disorders accounted for about 178 cases or 86.4% of all the errors [8].

AI can help in facilitating timely resuscitation and defibrillation. Machine learning, a component of AI, is an approach of exploring and developing algorithms that enable computers to learn and adapt to the various range of inputs from the ECG machine. Data sets arising from the ECG machines contain various types of data that pile up to form a large quantity of clinical information. In such a case, the systematic organization of these outputs becomes the key aspect for effective diagnosis. The algorithms implemented in this study helps us understand the hidden patterns in the ECG readings and automate the diagnostic process of Arrhythmia. In this process, machine learning algorithms help in the early stage identification of an Arrhythmia.

To address these concerns, In this paper, we implement machine learning algorithms to classify and predict Arrhythmia.

## 2 Methods

Our study uses the ECG dataset collected by Bilkent University, Turkey [9]. The data consists of 279 attributes and 452 observations, of which 44% are male, and 56% are female patients, with an average heart rate of 74.04bpm and 74.78bpm, respectively. The average age and weight of males are 45years/72.51kg and females 47years/63.82kg, respectively. The data has 16 classes (response variables) in which *Class 01* refers to 'normal' ECG. classes 02 to 15 refers to different classes of arrhythmia and *class 16* refers to the rest of unclassified ones. To better understand the data, we grouped the data into 12 sections based on their channel through which they were recorded. These 12 channels are named as D1-D3, V1-V6, AVR, AVL, and AVF. Each of these channels represents different waves (millisecond) and electric signals (millivolts) of the heart rhythm. Out of all the attributes, the first 9 are *age, sex, height, weight, average QRS duration (ms), the average duration between P and Q waves (ms), average duration between onset of Q and the offset of T waves (ms), average duration between two consecutive T waves (ms), and average duration of two consecutive P waves (ms)*. Other attributes in the data represent *vector angles* in degrees on the front plane of *QRS, T, P, QRST, and J*, respectively. The remaining measurements from 12 distinct channels are illustrated in Figure 1.

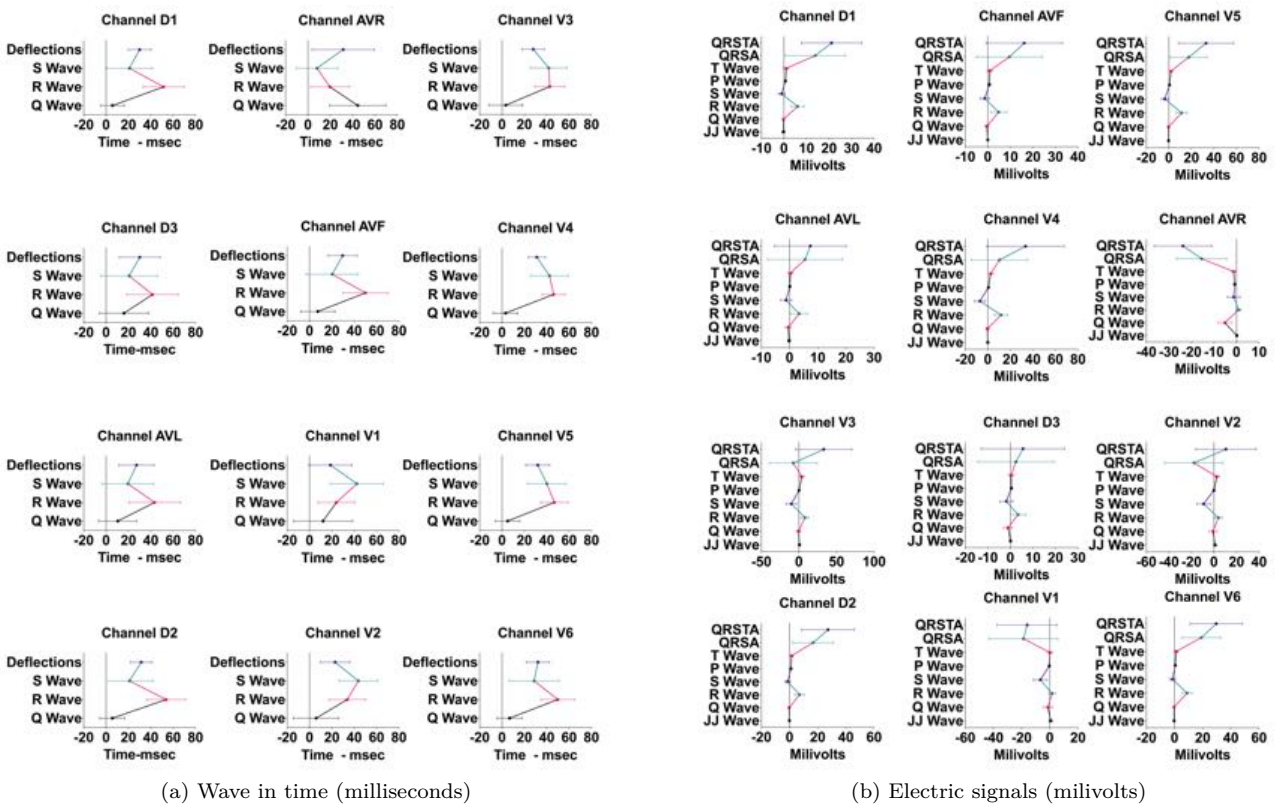


Figure 1: Information collected through 12 different channels

To simplify our model, we grouped the response variable into two sections, where 0 denotes the absence of arrhythmia, and 1 denotes arrhythmia. To minimize bias and variance in the dataset, we conducted a k-fold cross-validation technique on 80% of the raw data (training set). As illustrated in figure 2, in k-fold cross-validation, the original data sample is randomly partitioned into k proportional sub-samples with one sub-sample retained for model validation, and the remaining k-1 sub-samples used as training data. The cross-validation process is then reiterated k times, with the k results obtained from the k-folds averaged to produce a single estimation. Typical values for k range between 5 and 10 depending on study [10] in this analysis, we used k=5.

Typically, datasets are designated with several variables for practical model structure. Most of the predictor's relevance is unknown in advance. Several difficulties characterize the management of large datasets: (a) processing high dimensional datasets hinders computational speed; (b) excessive number of predictors (higher than optimal), impedes model accuracy. This problem, known as the minimal-optimal problem [11], has been intensively analyzed in the existing literature. Nevertheless, this pursuit of the practical consideration results in another problem, known as the all-relevant problem, which involves the identification of all attributes that

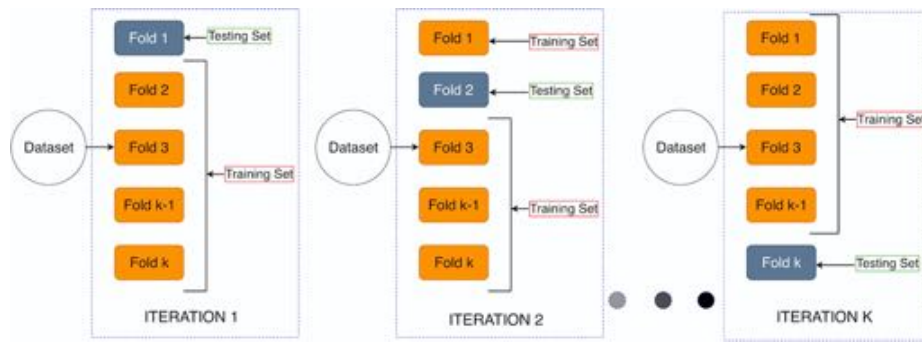


Figure 2: K-Fold cross validation

are germane for classification. However, deriving all relevant attributes may prove beneficial in understanding the fundamental mechanisms related to the subject of interest, rather than purely building a black box prognostic model. For example, when engaged in the classification of an arrhythmia dataset, the identification of all predictors related to the outcome (Healthy-0 or Arrhythmia-1) is necessary for a complete understanding of the process. In contrast, a minimal-optimal set of predictors (variables) might be more useful as classification markers. An honest discussion demarcating the importance of finding relevant attributes is detailed in [12, 11].

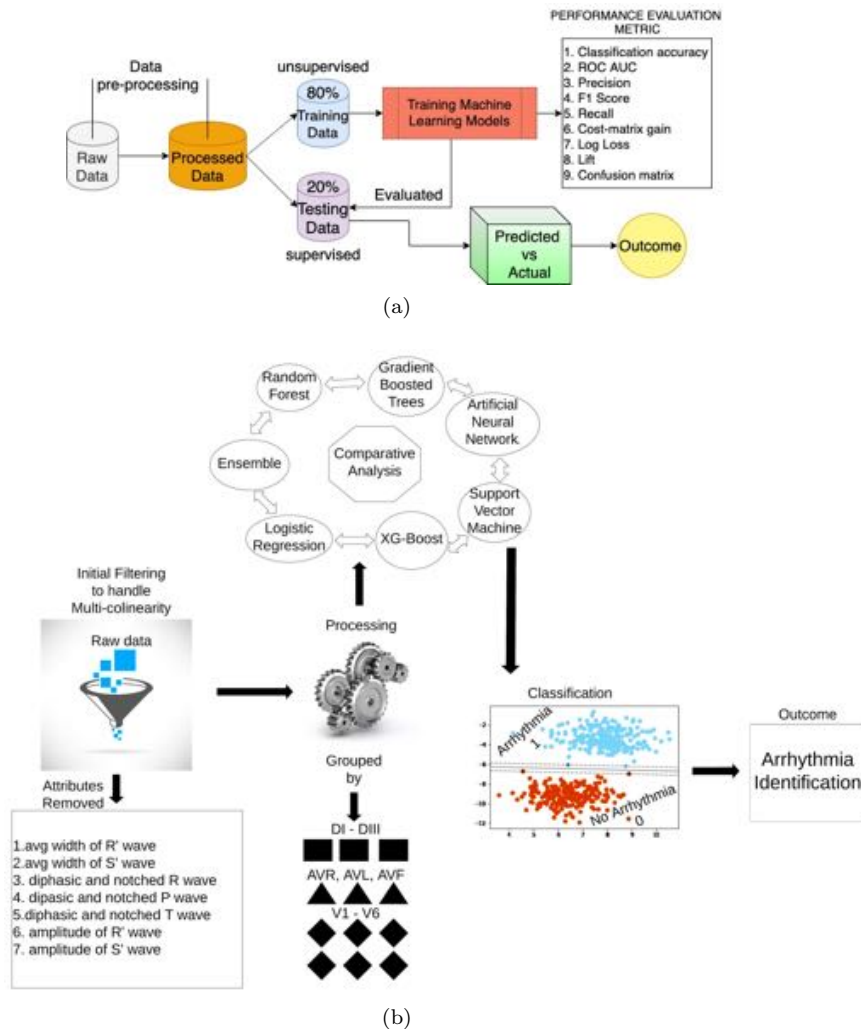


Figure 3: (a) Shows the overall framework of the method. (b) Illustrates the detailed steps taken to process the data, train different models and evaluate their classification performance

Our proposed framework consists of three main phases (a)pre-processing; (b)training; and (c)classification. The *pre-processing phase* primarily deals with the process of converting the raw data into a suitable format as required by the machine learning algorithms. In this phase, we removed attributes contributing to multicollinearity problems. We also handled the missing values, columns with more than 40% missing values were discarded from

further analysis. The *training phase* uses 80% of the pre-processed supervised patient’s dataset to train seven machine learning models. In the *prediction phase* of the framework, the trained model was used to predict the presence or absence of arrhythmia on a new dataset (20% test set). During this process, the performance of each model was evaluated primarily based on their *ROC AUC*, *classification accuracy*, and *false-negatives*. Other important evaluation metric is also recorded and has been discussed in the Result section.

Taking the cardiologist as a gold standard we implement seven machine learning algorithms: **(a)**Random Forest (RF), **(b)**Gradient Boosted Trees (GBT), **(c)**Logistic Regression (LR), **(d)**XG-Boost (XGB), **(e)**Artificial Neural Network (ANN), **(f)**Support Vector Machine (SVM), and **(g)** Ensemble of all. Explaining the machine learning models is not in the scope of this study.

All the model parameters were tuned to optimize the model’s performance for all analyses. All computations were performed using *Dataiku* and *Prism 8* for figure preparation. Figure 3 illustrates the method used in this study.

### 3 Results and discussion

In this study, the proposed Ensemble model and Gradient Boosted Trees with an area under the curve (AUC) of 0.87 and (0.87±0.08) respectively, outperformed other models with an AUC of 0.85(RF), 0.80 (LR), 0.86 (XGB), 0.78 (ANN), and 0.84 (SVM). Additionally, ensemble, logistic regression, and support vector machines outperformed the cardiologist’s (recall 0.78) performance in arrhythmia diagnosis [13] and also outperforms the exiting VF15 accuracy of 62% by more than 20% margin. Our model, Ensemble (recall 0.86; F1 Score 0.81) and support vector machine (recall 0.91) also outperformed deep learning method (recall 0.78; F1 Score 0.776) developed by Rajpurkars et al., in 2017 [5]. Table 1 shows the performance of all the machine learning algorithms tested in this study. As reported in table 1, Ensemble model yielded the highest classification accuracy (0.84), F1 Score (0.81), Cost-matrix gain (0.31), and ROC (0.87). Following closely, with a comparable performance, Gradient Boosted Trees yielded the second-best classification accuracy of 0.81±0.13, highest precision (0.77±0.13), ROC (0.87±0.08) and highest lift (1.95±0.52). ROC AUC of Random Forest and XG-Boost were also near Ensemble and GBT. As a graphical illustration, the performance and decision cut-off (threshold) of the Ensemble model is also shown in figure 4. Figure 5 illustrates the AUC for all k-fold for each machine learning model.

Despite having a high-performance outcome, the Ensemble model fails to identify 14% of Arrhythmia, resulting in false-negative (Table 2). A false-negative outcome in the healthcare industry can delay proper diagnosis. Delayed or misdiagnosis can not only increase treatment cost but also can worsen patient health, safety, and anxiety leading to undesirable consequences. Although the support vector machine performed poorly than the other models, it has the least false-negative outcome of 11%.

Unlike false-negatives, false-positives have no deterring impact on patient health. However, it contributes significantly to the diagnostic expenses and patient’s (patient family) anxiety and stress levels. The logistic regression model in this study yielded the highest false-positive outcome (40%), followed by the support vector machine (33%).

Table 1: Performance of all machine learning algorithm used in this study

Model	Accuracy	Precision	Recall	F1 Score	Gain	Log Loss	ROC	Lift
Ensemble	<b>0.84</b>	0.77	0.86	<b>0.81</b>	<b>0.31</b>	0.45	<b>0.87</b>	1.90
RF	0.78±0.06	0.72±0.13	0.72±0.12	0.72±0.08	0.18±0.24	0.51±0.04	0.85±0.06	1.86±0.27
GBT	0.81±0.13	<b>0.77±0.13</b>	0.76±0.21	0.76±0.16	0.24±0.09	0.46(0.24)	<b>0.87(0.08)</b>	<b>1.95±0.52</b>
LR	0.66±0.06	0.54±0.10	0.83±0.08	0.66±0.08	0.20±0.15	0.54(0.07)	0.80(0.08)	1.69±0.17
XGB	0.80±0.11	0.74±0.16	0.77±0.21	0.75±0.13	0.17±0.25	0.55(0.07)	0.86(0.08)	1.92±0.0.31
ANN	0.75±0.06	0.70±0.17	0.65±0.08	0.67±0.09	0.20±0.07	<b>0.71±0.22</b>	0.78±0.06	1.66±0.24
SVM	0.71±0.12	0.59±0.16	<b>0.91±0.07</b>	0.71±0.13	0.21±0.18	0.48±0.03	0.84±0.03	1.79±0.17

Table 2: Confusion matrix

		Ensemble		RF		GBT		LR		XGB		ANN		SVM	
		Predicted - Arrhythmia													
		1	0	1	0	1	0	1	0	1	0	1	0	1	0
Actual - Arrhythmia	1	86%	14%	75%	25%	69%	31%	83%	17%	72%	28%	58%	42%	<b>89%</b>	11%
	0	17%	83%	17%	83%	15%	85%	40%	60%	15%	85%	12%	<b>87%</b>	33%	67%

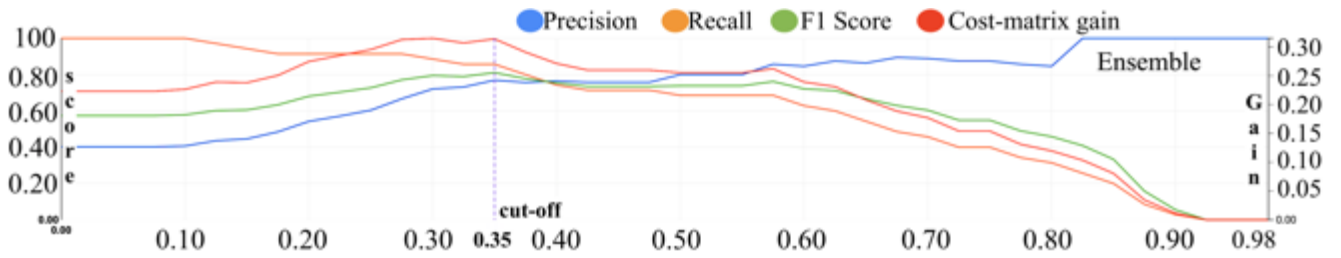
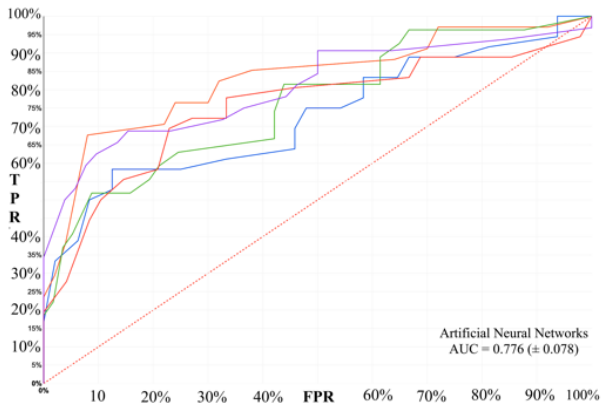
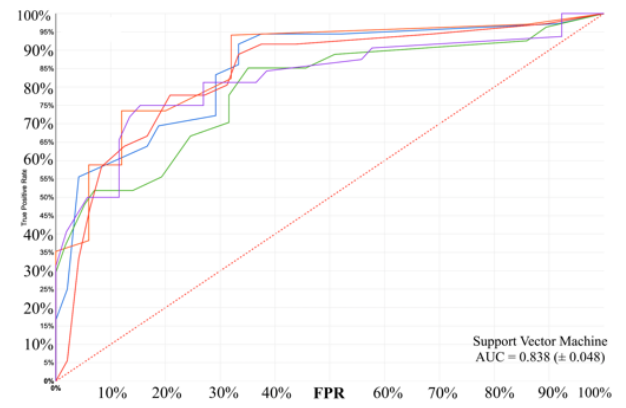


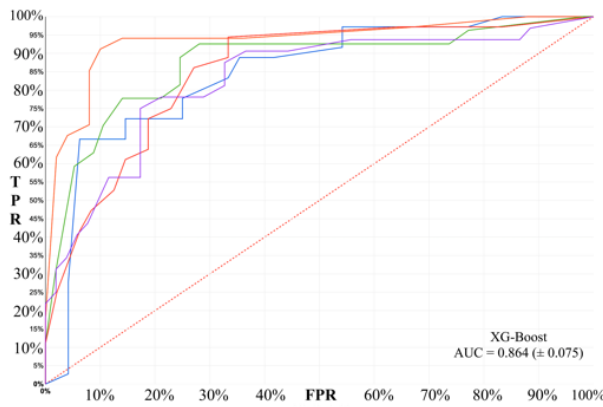
Figure 4: Decision cut-off for ensemble model



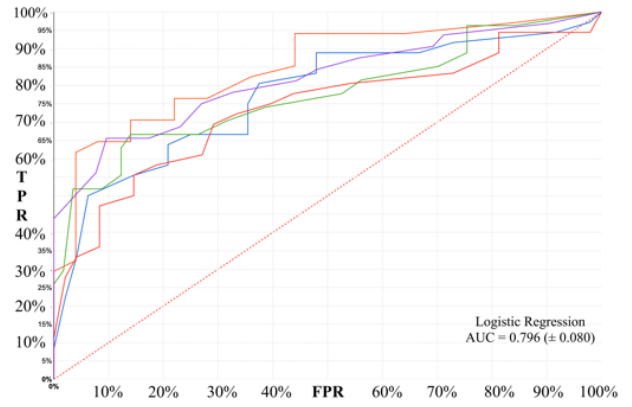
(a) ROC-AUC (ANN)



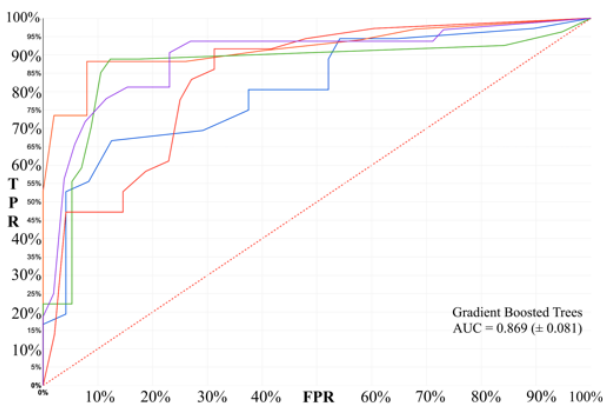
(b) ROC-AUC (SVM)



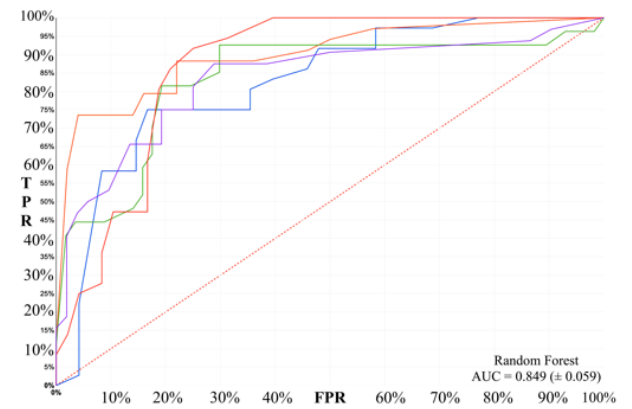
(c) ROC-AUC (XGB)



(d) ROC-AUC (LR)



(e) ROC-AUC (GBT)



(f) ROC-AUC (RF)

Figure 5: Area under ROC. Each colored curve represents the ROC for 5 different folds (k-fold cross validation). The annotated AUC in each facet of the figure (a)-(f) is the mean AUC  $\pm$  its standard deviation

## 4 Conclusion

In this paper, we illustrate the use of machine learning techniques in diagnosing Arrhythmia. Additionally, with the help of a comparative analysis of seven machine learning techniques, this study highlights the strengths and weaknesses of several algorithms. This study acknowledges the advantages of using a machine learning technique by showcasing its predictive capabilities, reliability, and accuracy over cardiologists and other existing methods. The potential of machine learning to foster timely and precise diagnosis can contribute to better care and improve patient safety.

As precautionary measures, while implementing machine learning techniques, we recommend referring to false-negative outcomes of each model. It can prevent not only fatal outcomes (misdiagnosis can lead to death) but also prevent delayed treatment. The proposed framework is not generic and might not produce a similar result when applied on a different dataset. To further improve the robustness and performance of the models proposed in this study, future work should include better data that represent a broader population and evaluate other algorithms.

## References

- [1] Q. Qin, J. Li, L. Zhang, Y. Yue, and C. Liu, “Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification,” *Scientific reports*, vol. 7, no. 1, p. 6067, 2017.
- [2] R. E. Lane, M. Cowie, and A. W. C. Chow, “Prediction and prevention of sudden cardiac death in heart failure,” *Heart*, vol. 91, no. 5, pp. 674–680, 2005.
- [3] R. Deo and C. M. Albert, “Epidemiology and genetics of sudden cardiac death,” *Circulation*, vol. 125, no. 4, pp. 620–637, 2012.
- [4] N. Srinivasan and R. Schilling, “Sudden cardiac death and arrhythmias,” *Arrhythmia & electrophysiology review*, vol. 7, no. 2, p. 111, 2018.
- [5] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, “Cardiologist-level arrhythmia detection with convolutional neural networks,” *arXiv preprint arXiv:1707.01836*, 2017.
- [6] G. Lopera and A. B. Curtis, “Risk stratification for sudden cardiac death: current approaches and predictive value,” *Current cardiology reviews*, vol. 5, no. 1, pp. 56–64, 2009.
- [7] S. Raj and K. C. Ray, “A personalized arrhythmia monitoring platform,” *Scientific reports*, vol. 8, 2018.
- [8] M. Guglin and D. Thatai, “Common errors in computer electrocardiogram interpretation,” *International journal of cardiology*, vol. 106, no. 2, pp. 232–237, 2006.
- [9] H. A. Guvenir, B. Acar, G. Demiroz, and A. Cekin, “A supervised machine learning algorithm for arrhythmia analysis,” in *Computers in Cardiology 1997*, pp. 433–436, Sep. 1997.
- [10] M. Kuhn and K. Johnson, *Applied predictive modeling*, vol. 26. Springer, 2013.
- [11] R. Nilsson, J. M. Peña, J. Björkegren, and J. Tegnér, “Consistent feature selection for pattern recognition in polynomial time,” *Journal of Machine Learning Research*, vol. 8, no. Mar, pp. 589–612, 2007.
- [12] A. Choudhury and B. Eksioglu, “Using predictive analytics for cancer identification,” in *Proceedings of the 2019 IISE Annual Conference. Edited by H.E. Romeijn AS, R. Thomas.*, pp. 1–6, May 2019.
- [13] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nature medicine*, vol. 25, no. 1, p. 65, 2019.