Supervised Machine Learning Based Cardiac Arrhythmia Analysis and Detection

Riyadul Islam[1], Md. Lutfor Rahman[1], Masfik Alam[1]

[1]Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka-1229, Bangladesh
[*]Corresponding Author: Mohammad Monirujjaman Khan
Email: monirujjaman.khan@northsouth.edu

Abstract: Cardiac arrhythmia, often commonly known as a serious type of heart arrythmia, encompasses a variety of arrythmias that affect the heart and has been the leading cause of mortality globally in recent decades. It is associated with numerous risks in heart arrythmia and a requirement of the moment to get accurate, trustworthy, and reasonable methods to establish an early diagnosis in order to accomplish early arrythmia treatment. In the healthcare sector, data analysis is a widely utilized method for processing massive amounts of data. Researchers use a variety of statistical and machine learning methods to evaluate massive amounts of complicated medical data, assisting healthcare practitioners in predicting cardiac arrythmia. This study covers many aspects of cardiac arrythmia, as well as a model based on supervised learning techniques such as Random Forest, Decision Tree, and Logistic Regression. The purpose of this study article is to forecast the likelihood of individuals getting heart arrythmia. The findings show that Logistic regression achieves the best accuracy score (80.10 percent).

Keywords: Cardiac arrythmia, data analysis, machine learning, UCI database, Logistic regression

1 Introduction

It is difficult to diagnose cardiac arrythmia due to the presence of many health problems such as diabetes, high blood pressure, excessive cholesterol, and an irregular pulse rate. Numerous data analysis and neural network methods have been used to determine the severity of cardiac arrythmia in people. The severity of illness is categorized using a variety of techniques, including the K-Nearest Neighbor (KNN) algorithm, Decision Trees (DT), Genetic Algorithm (GA), and Naive Bayes (NB) algorithm [1-2]. Due to the complexity of cardiac arrythmia, it must be treated with caution. Failure to do so may have a detrimental effect on the heart or result in early death. Medical science and statistical perspectives are utilized to identify different types of metabolic disorders. Data analysis with categorization is essential for heart arrythmia prediction and data research. Additionally, we have seen decision trees used to predict the accuracy of heart arrythmia-related events [3]. Numerous approaches for knowledge abstraction have been employed in conjunction with well-established data mining techniques for the heart arrythmia diagnosis. Numerous analyses have been conducted in this study to develop a prediction model, not only utilizing different methods but also by connecting two or more techniques. Data mining is the process of extracting needed information from massive databases in a variety of areas, including medicine, business, and education. Machine learning is one of the fields of artificial intelligence that is advancing at a breakneck pace. These algorithms are capable of analyzing massive amounts of data from a variety of areas, one of which being the medical field. It is a replacement for the conventional prediction modeling method that use a computer to acquire a knowledge of complicated and non-linear interactions between many variables by minimizing the difference between anticipated and actual results [4]. Data mining is the process of sifting through massive datasets in order to extract critical decision-making information from a collection of historical records for future study. The medical profession is replete with patient data. These data must be analyzed using a variety of machine learning techniques. Healthcare experts analyze this data in order to make appropriate diagnostic decisions. Through analysis, medical data mining using classification algorithms offers therapeutic assistance. It evaluates methods for classifying patients' risk of developing heart arrythmia [5].

Several studies have been performed, and numerous machine-learning models have been deployed, all with the goal of classifying and forecasting heart arrythmia diagnosis. ANNs were developed to achieve the greatest prediction accuracy possible in the medical sector [6]. ANNs are used to forecast cardiac arrythmia via back propagation multilayer perception (MLP). The resulting findings are compared to those of previously published models in the same area and found to be significantly improved [7]. The UCI laboratory's data on heart arrythmia patients is utilized to identify patterns using NN, DT, Support Vector Machines SVM, and Naive Bayes. The performance and accuracy of various algorithms are compared. The suggested hybrid approach achieves an accuracy of 86.8 percent for F-measure, which is comparable to the other available methods [8]. The classification of Convolutional Neural Networks (CNN) without segmentation is presented. This technique considers cardiac cycles with a variety of start locations derived from Electrocardiogram (ECG) data during the training phase. CNN is capable of generating features with varying locations throughout the patient's testing phase [9], [10]. Previously, a significant quantity of data produced by the medical sector was not used properly. The novel methods described here reduce the cost and enhance the accuracy of heart arrythmia prediction in a simple and efficient manner. The numerous research approaches examined in this study for prediction and classification of heart arrythmia utilizing machine learning (ML) and deep learning (DL) techniques are very accurate in proving these methods' effectiveness [11], [12]. Avinash Golande and colleagues investigate a variety of machine learning methods that may be used to classify cardiac arrythmia. A research was conducted to evaluate the accuracy of Decision Tree, KNN, and K-Means algorithms that may be utilized for classification [13]. This study indicates that Decision Trees achieve the greatest accuracy, and that they may be made more efficient via the use of a mix of various methods and parameter tweaking. T.Nagamani and colleagues [14] developed a system that combined data mining methods with the MapReduce algorithm. For the 45 instances in the testing set, the accuracy achieved in this study was higher than the accuracy obtained using a typical fuzzy artificial neural network. Due to the usage of dynamic schema and linear scaling, the accuracy of the method was increased in this case. Fahd Saleh Alotaibi developed a machine learning model that compares five distinct methods [15]. Rapid Miner was employed, which provided a better level of accuracy compared to Matlab and Weka. In this study, the classification algorithms Decision Tree, Logistic Regression, Random Forest, Naive Bayes, and SVM were compared for accuracy. The decision tree algorithm was the most precise. Anjan Nikhil Repaka, et al., developed a system in [16] that combines NB (Naive Bayesian) methods for dataset categorization and AES (Advanced Encryption Standard) for secure data transmission for illness prediction. Theresa Princy, R., et al. conducted a study comparing several categorization algorithms used for heart arrythmia prediction. The classification methods utilized were Naive Bayes, KNN (Kindest Neighbour), Decision tree, and Neural network, and the accuracy of the classifiers was evaluated over a range of attribute counts [17]. Nagaraj M Lutimath et al. used Naive Bayes classification and SVM to predict cardiac arrythmia (Support Vector Machine). The performance metrics utilized in the study are the Mean Absolute Error, the Sum of Squared Error, and the Root Mean Squared Error. It has been shown that SVM outperforms Naive Bayes in terms of accuracy [18].

2 Methodology and Method

This section includes details on the methods and materials utilized, as well as a dataset description, a schematic diagram, machine learning algorithms, and evaluation matrices.

2.1 Dataset

The Heart Arrythmia Dataset was utilized, which is a compilation of four distinct databases, however only the UCI Cleveland dataset has been used [19]. This database has 76 characteristics in total, however all published studies use a subset of just 14 features [20]. As a result, for our study, we utilized the previously processed UCI Cleveland dataset accessible on the Kaggle website. Table 1 below has a detailed explanation of the 14 characteristics utilized in the proposed study.

2.2 Schematic Diagram of The System

The proposed study indicated heart arrythmia by examining the three classification methods listed above and doing performance analysis. The goal of this research is to accurately predict whether or not the patient has heart arrythmia. The input values from the patient's health report are entered by the health professional. The data is incorporated into a model that forecasts the chance of developing heart arrythmia. Figure 1 depicts the full procedure.

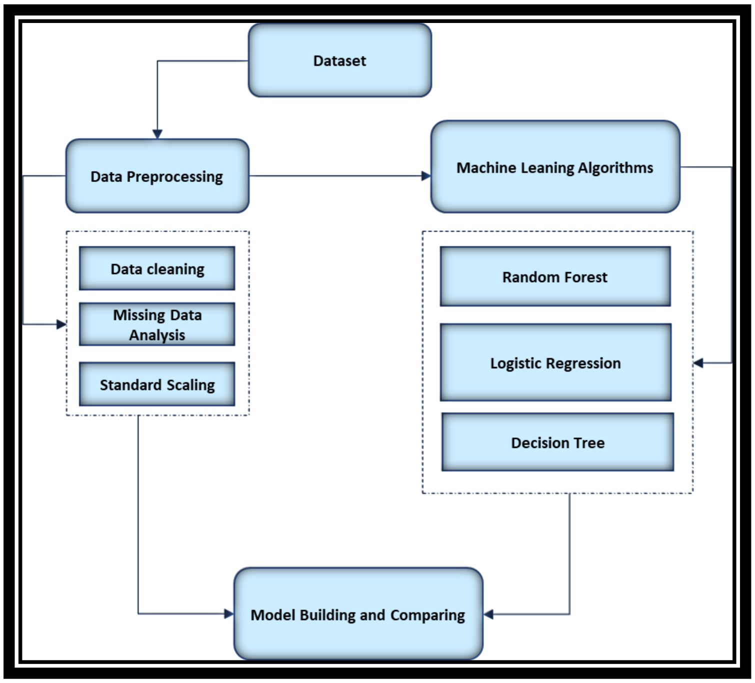

Figure 1: System's Schematic Diagram

The input dataset is divided into 80% of the training dataset and 20% of the test dataset. A training dataset is a collection of data that is being used to train a model. The testing dataset is also used to evaluate the trained model's performance. The performance of each method is generated and analyzed using a variety of measures, including accuracy, precision, recall, and F1-scores, as discussed below.

2.3 Machine Learning Algorithms

Classification and regression techniques based on Random Forests are utilized. It constructs a tree for the data and then makes predictions using that tree. The Random Forest technique is capable of processing enormous datasets and producing the same result even when substantial portions of the record values are missing. The decision tree's produced samples may be stored and used on additional data. There are two steps in random forest: first, generate a random forest, and secondly, using the random forest classifier built in the first stage, make a prediction. Figure 2 shows the schematic diagram of random forest algorithm.

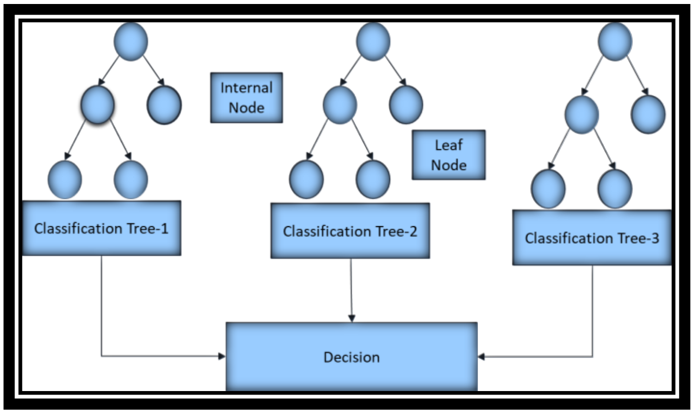

Figure 2: Schematic Diagram of Random Forest

Also, Random Forest is a Decision Tree-based method. After combining numerous separate decision trees, it is generally more accurate and reliable than a single tree. The random selection of samples and features, as well as the integration procedures, give a Random Forest an edge over a Decision Tree. While the former resists overfitting better, the latter is more accurate. Random Forest uses the Decision Tree as the bagging model. The Decision Tree method is represented visually as a flowchart, with the central node representing the dataset's properties and the outside branches representing the result. Decision Trees are selected because they are quick, dependable, simple to read, and need little data preparation. In a Decision Tree, the prediction of the class label begins at the tree's root. The root attribute's value is compared to the record's attribute. The matching branch is explored to that value and a move is made to another node based on the outcome of the comparison. Figure 3 shows the schematic diagram of decision tree algorithm.

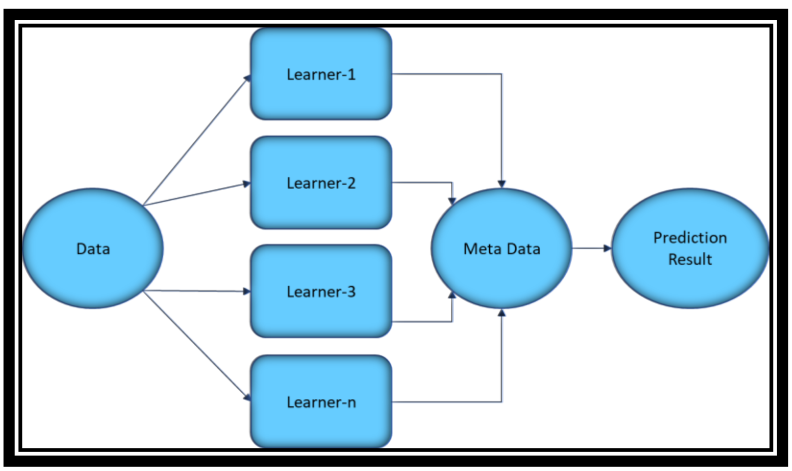

Figure 3: Schematic Diagram of Decision Tree

Strategic splits have a big influence on a decision tree's accuracy. They use different decision criteria. The development of sub-nodes increases their homogeneity. Because the target variable grows, the node's purity rises. A decision tree is simple to grasp and can handle both numerical and categorical data. Linear Regression is a statistical approach that is often used to solve issues involving binary classification. Rather of fitting a straight line or hyper plane, logistic regression employs the logistic function to constrain the output of a linear equation to the range 0 to 1. Due to the presence of 13 independent variables, logistic regression is well-suited for categorization. Figure 7 shows the schematic diagram of logistic regression algorithm.

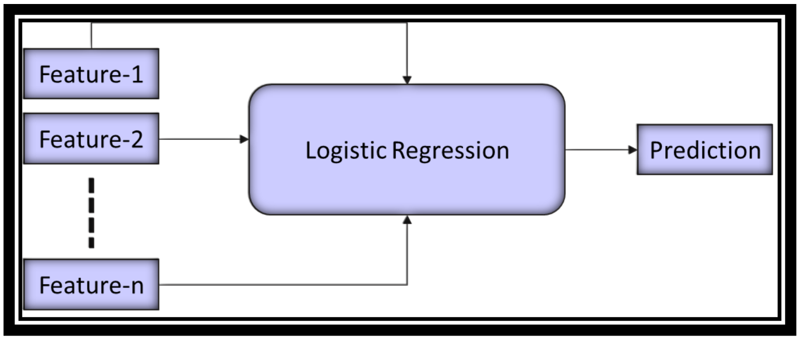

Figure 3: Schematic Diagram of Logistic Regression

2.4 Block Diagram of Confusion Matrix

A confusion matrix is a technique for describing the performance of a classification system. The number of correct and incorrect predictions is summed and denoted by count values. This is the key to the misunderstanding matrix. The system's performance block diagram is shown in Figure 4.

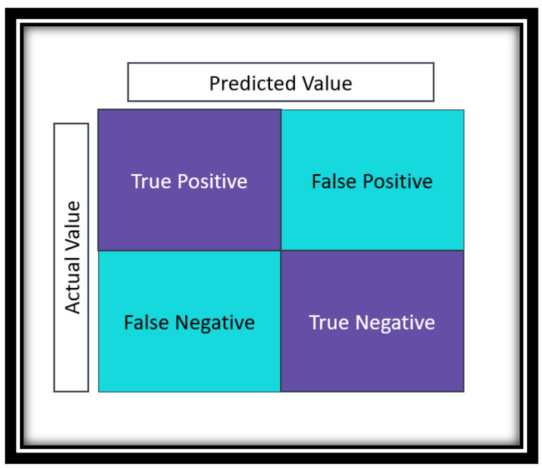

Figure 4: Block Diagram of Confusion Matrix

It elucidates not only the errors made by the classifier, but also the kind of faults committed. The expected row and predicted column for a class include the total number of correct predictions. Similarly, the expected row and projected column for a class value include the total number of incorrect guesses.

3 Result and Data Analysis

This section discusses the capabilities of the models, model predictions, inquiry, and final outcomes.

3.1 Model Accuracy

Table 1 shows the three different model's classification result.

Table 1: Classification result of three different models

Model	Precision	Recall	F1-score	Accuracy

Random Forest	77%	87%	82%	80%
Decision Tree	71%	74%	72%	72%
Logistic Regression	**88%**	**88%**	**88%**	**88%**

According to Table 1, Logistic Regression outperformed the other algorithms in terms of accuracy. The Random Forest also performed well in terms of accuracy. The performance of the Decision Tree, on the other hand, is really low.

4 Conclusion

Three machine learning techniques were provided in this work, and their comparative assessment is described. The goal of the article was to determine which machine learning classifier would be the most effective in predicting heart arrythmia based on the dataset utilized. Three classifiers were built and their results were compared. Some of the comparison approaches used include the confusion matrix, accuracy, specificity, and sensitivity. For the 14 variables in the sample, the Logistic Regression classifier performed admirably in the ML approach. The Logistic Regression technique outperformed the other two classifiers employed, with an accuracy of 88 percent. The Random Forest classifier had an accuracy of 80%, whereas the Decision Tree classifier had an accuracy of 72%. This idea has the potential to be a game changer in the medical field. Patients at risk of heart arrythmia might be recognized quickly with this method, which could help to lower the rising death rate. The properties in the dataset that the prediction model is built on are not prohibitively costly to record. As a result, this kind of diagnostic may be made accessible to patients at a reasonable cost, allowing it to reach a considerably larger number of people. This kind of diagnosis will become more common in the future as machine learning algorithms improve as a result of continuous research. If additional patient information is utilized, the model may be refined and adjusted. A bigger dataset ensures more precise and accurate findings. This is critical since medical diagnosis is a very delicate problem that requires high degrees of accuracy and precision.

Acknowledgement:

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Durairaj. V.Revathi, 2015. Prediction of Heart Arrythmia Using Back Propagation MLP Algorithm, 4(08), pp.235−239.

[2] M.Gandhi, 2015. Predictions in Heart Arrythmia Using Techniques of Data Mining. 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp.520−525.

[3] A.S. Abdullah, 2012. A Data mining Model for predicting the Coronary Heart Arrythmia using Random Forest Classifier, (Icon3c), pp.22 −25.

[4] S.F.Weng, J.Reps, J. Kai, J.M.Garibaldi, N.Qureshi. Can machinelearning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.

[5] V.V. Ramalingam, A.Dandapath, M.K. Raja. Heart arrythmia prediction using machine learning techniques: a survey. Int J Eng Technol. 2018;7(2.8):684–7.

[6] Baccour, L., 2018. Amende d fuse d TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets R. Expert Systems With Applications, 99, pp.115−125. Available at: https://doi.org/10.1016/j.eswa.2018.01.025.

[7] Das, R., Turkoglu, I. Sengur, A., 2009. Expert Systems with Applications Effective diagnosis of heart arrythmia through neural networks ensembles. Expert Systems with Applications, 36(4), pp.7675−7680. Available at: http://dx.doi.org/10.1016/j.eswa.2008.09.013.

[8] Cheng, C. Chiu, H., 2017. An Artificial Neural Network Model for the Evaluation of Carotid Artery Stenting Prognosis Using a NationalWide Database. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp.2566−2569.

[9] Nahar, J., Imam, T., Tickle, K.S., et al., 2013. Expert Systems with Applications Association rule mining to detect factors which contribute to heart arrythmia in males and females. Expert Systems with Applications, 40(4), pp.1086−1093. Available at: http://dx.doi.org/10.1016/j.eswa.2012.08.028.

[10] Zaman, S., 2017. Codon Based Back Propagation Neural Network Approach to Classify Hypertension Gene Sequences, pp.443−44

[11] Ravish, D.K. Shenoy, N.R., 2014. Heart Function Monitoring , Prediction and Prevention of Heart Attacks: Using Artificial Neural Networks. , pp.1−6.

[12] Zhang, W. Han, J., 2017. Towards Heart Sound Classification Without Segmentation Using Convolutional Neural Network, 44, pp.1-4.

[13] Avinash Golande, Pavan Kumar T, 'Heart Arrythmia Prediction Using Effective Machine Learning Techniques', International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[14] T.Nagamani, S.Logeswari, B.Gomathy, 'Heart Arrythmia Prediction using Data Mining with Mapreduce Algorithm', International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[15] Fahd Saleh Alotaibi, 'Implementation of Machine Learning Model to Predict Heart Failure Arrythmia', (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[16] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, 'Design And Implementation Heart Arrythmia Prediction Using Naives Bayesian', International Conference on Trends in Electronics and Information(ICOEI 2019).

[17] Theresa Princy R,J. Thomas,'Human heart Arrythmia Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[18] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,'Prediction Of Heart Arrythmia using Machine Learning', International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[19] Heart Arrythmia UCI. Available at: https://www.kaggle.com/datasets/bulentesen/cardiac-arrhythmia-database

[20] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, ―Medical Data Mining for Heart Arrythmias and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.