Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766192

THE REAL TIME DETECTION OF CARDIAC ARREST USING MACHINE LEARNING TECHNIQUES

[1]N Srikanth Reddy, [2]K Sai Vatsav, [3]K Rajeswari, [4*]Dr P M Manohar
[1,2,3] 4th B.Tech Students, Dept. of Computer Science and Engineering, Raghu Engineering College, Visakhapatnam
[4]Associate Professor, Dept. of Computer Science and Engineering, Raghu Engineering College, Visakhapatnam
[1]19981a05a8@raghuenggcollege.in, [2]19981a0582@raghuenggcollege.in ,
[3]19981a0570@raghuenggcollege.in, [4]manohar.pattisapu@raghuenggcollege.in,

Abstract:
Cardiac arrest occurs more often in younger people. Those who are dying at a younger age these days due to health instability should seek immediate medical attention. When the heart breaks down and abruptly stops beating, it can cause an irregular pulse, which can lead to cardiac arrest. Rapid loss of awareness, cessation of respiration, or scarcely audible gasps occur. Death happens in a few minutes. There are a tonne of cardiac arrest datasets available in the healthcare sector. By collecting real-time data from sensors and feeding it into a prediction model, it is possible to detect cardiac arrests with a low-to-high probability. To make model predictions, machine learning algorithms are being used. In this present circumstance, we employ six particular machine learning algorithms: decision tree, logistic regression, random forest classifier, support vector machine, and K-nearest neighbour. The most effective of these methods is the decision tree and random forest classification. The model's forecast accuracy is 100 percent. Given that it has a number of advantages over the other, the random forest classifier is the best among the two finest algorithms. Finally, the proposed model is used to predict whether or not a cardiac arrest will occur within the next ten years, while also being used outside of the hospital. This will help alert the patient to consult the doctor.

Keywords: Cardiac arrest, machine learning, smart watches, sphygmomanometer, and glucometer.

1. Introduction

A sudden and unexpected loss of heart function is referred to as cardiac arrest. This can happen when the electrical system of the heart malfunctions, preventing the heart from moving blood. As a result, the brain and other vital organs are deprived of oxygen-rich blood, which can lead to serious and potentially fatal consequences within minutes. Cardiac arrest can occur in people of all ages, although it is more common in older adults and those with underlying heart conditions. A variety of factors can contribute to it, including injuries, cardiac illness, electrolyte imbalances, substance overdoses, heart attacks, and other medical issues[15]. The early detection and treatment of cardiac arrest are critical to survival. Cardiopulmonary resuscitation (CPR) and defibrillation, which delivers an electric shock to the heart, are the two main interventions used to restore normal heart rhythm and circulation[15][16]. Early administration of these interventions can greatly increase the chances of survival and minimise the chance of cerebral injury. Notwithstanding the way that the terms are periodically utilised reciprocally, it is critical to comprehend that heart failure and a coronary episode are two distinct things. Cardiovascular failures happen when the heart's blood supply is cut off; heart failure, notwithstanding, happens when the heart's electrochemical instrument glitches[16].

Cardiac arrest is a major contributor to morbidity and mortality worldwide. Addressing this issue requires a comprehensive public health approach, including efforts to prevent cardiac arrest, improve bystander response and emergency medical services, and advance clinical care and research[6]. Cardiac arrest is associated with significant economic costs, including healthcare expenditures, lost productivity, and disability. The cost of treating cardiac arrest and its associated complications can be substantial, making prevention and early intervention critical. Cardiac arrest is associated with high mortality rates[17]. Without prompt intervention, the chances of survival decrease rapidly. Survival rates vary depending on the location and timing of the cardiac arrest, as well as the effectiveness of interventions such as CPR and defibrillation[17][18]. Overall, the study of cardiac arrest is critical for improving survival rates, preventing the condition, and developing more effective treatments to address this significant public health issue. The motivation behind this study is to improve the quality of care for patients who suffer from cardiac arrest. By understanding the causes, risk factors, prevention, and treatment of cardiac arrest, healthcare professionals can develop better treatment protocols and improve patient outcomes. The real-time data collected from the sensors is then used to detect whether cardiac arrest has occurred or not. Without going to the hospital, it will help whether we have to go to the hospital or not.

Problem Statement

Cardiac arrest is a leading cause of death worldwide. To increase awareness of cardiac arrest and improve access to emergency medical services, ultimately leading to improved outcomes for peoples.

Contribution

To increase awareness of cardiac arrest into peoples our propose system is using the real time detection sensors collecting the data passing to our machine learning model than it alert the people may or not occurred cardiac arrest this will help the people without going hospital it suggest the whether the peoples need to consult doctor or not.

On the Kaggle website, the dataset for cardiac arrest is accessible to everyone. It contains the data of over 4,040 inhabitants of the Massachusetts town of Framingham. What it means to use this dataset is to determine if the patients will suffer at least a CHD (coronary heart disease) or not in ten years. To be able to reach this goal, for each person there are fifteen attributes divided into three groups: demographic, medical history of the patient, and his or her current medical status (such as BMI, heart rate, etc.). Each characteristic has the potential to predict whether the patient will experience a CHD in ten years, and various research has shown predictions based on datasets that are freely available. Nevertheless, this study's objective is to identify Based on information obtained in real time from people using sensors and equipment, an early risk prediction of cardiac arrest is made. In order to identify which model performs the best, we develop a model in this post utilising ML techniques and pertinent datasets. The points raised The following are the primary contributions of the work:

1. Using sensors and equipment, collect an individual's vital signs in real time.
2. Using the ML algorithm to detect cardiac arrest
3. To present a comparison of various ML classifiers.
4. Real-time data collection and prediction is presented through the web-based user interface.

The remaining article will follow the literature review, it shows previous years research data of cardiac arrest and then follows the methods in this the proposed flow diagram that shows various stages need to implement, then they are interface (UI), prediction, model evaluation, algorithms, training and test data, data preprocessing and the use of CSV files, performing the exploratory data analysis, scaling and evaluation metric, then finally concludes the experimental results, conclusion and future work.

2. Literature Review

This section talks about how various methods and datasets with different features and classification algorithms can be used to predict and diagnose heart problems.

Minsu Chae et al. describe early cardiac arrest detection and speedy emergency response in high-risk individuals. Patients who were hospitalised at the Soonchunhyang College Cheonan Medical Clinic between January 2016 and June 2019 made up the examination populace. Patients younger than 18 and those who kicked the bucket or went into heart failure in no less than eight hours after appearance are excluded from this review. Here, only adults above the age of 18 made the forecast. The proposed Long Short-Term Memory (LSTM) model has an 85.92% positive predictive value and an 89.70% sensitivity [1].

Youngam Lee et al. A profound learning-based approach for perceiving people in heart failure demonstrated high responsiveness and a low deception rate in a multicenter preliminary study. From June 2010 to July 2017, patients who were owned by up to two emergency clinics were analysed in this associate review. There were 52,131 patients canvassed in total. Information from February to July 2017 was utilised to test the result. They have made an early warning framework in view of profound learning (DEWS). This increased sensitivity by up to 24.3% and decreased alerts by 41.6% [2].

Erik Alonso et al. In the event that out-of-hospital heart failure could be treated using automatic pulse detection (OHCA). The Tualatin Valley Fire and Salvage in Tigard, Oregon, utilised the Philips Heart Start MRx screen/defibrillator (Philips Medical Services, Mama, USA) to treat 187 OHCA patients somewhere in the range of 2010 and 2014. Signals for ECG, TI, and capnography were available in each episode. The sensitivity, specificity, balanced accuracy, and total accuracy mean (standard deviation) values of the five-feature classifier that comprised the SVM optimal solution were 92.4 (0.7), 93.7 (0.8), and 92.6 (0.5), respectively [3].

Joon-myoung Kwon et al. Cardiac arrests in hospitals place a heavy budgetary burden on the system. An emergency clinic provided 32,294 ECGs from 10,461 patients as improvement information and 4483 ECGs from 4483 patients as inward approval information, individually. As extra outer approval information, 10,728 ECGs from 10,728 patients from another emergency clinic were used. While anticipating heart failure in no less than 24 hours, the DLA's regions under the beneficiary's working trademark bends were 0.913 and 0.948, respectively [4].

L. Murukesh et al. The destructive coronary illness known as abrupt heart failure claims a great many lives. (SCA) . The proposed signal handling innovation has yielded critical outcomes that look to anticipate SCA two minutes ahead of time. Two global norms datasets, as well as the MIT/BIH Unexpected Heart Passing (SCD) Holter Information Base for SCA expectations. Here, the ECG information assortment is utilised. Support Vector Machine (SVM) and Probabilistic Brain Organisation (PNN) were utilised to foresee the SCA and ordinary control circumstances. SVM and PNN have the most extreme mean SCA estimate paces of 96.36% and 93.64%, respectively [5].

Apeksha Shah et al. The medical services industry has a sizable information base of datasets for study and determination purposes. Sensors were utilised to assemble continuous information, which was then saved on a distributed computing device like Google Firebase. The likelihood of heart failure is anticipated to be identified utilising six AI calculations: the fake brain organisation (ANN), the arbitrary timberland classifier (RFC), the angle help outrageous inclination supporting (XGBoost) classifier, the help vector machine (SVM), the guileless Bayes (NB), and the choice tree (DT). With a general precision of 98%[6].

3. Methods

The process of building a machine learning model involves several key components, including the user interface (UI), prediction, model evaluation, algorithms, training and test data, data pre processing, and the use of CSV files. To start, the model is created using algorithms and trained on a set of data that includes both the training and testing sets. Before training the model, the data must be prep processed to remove any inconsistencies and ensure it is ready for the model to learn from. The data can come from various sources, such as CSV files. Once the model is trained, it is evaluated using the test data that it has not seen before, and the accuracy of the model is measured based on its performance. The user interface is an essential part of the machine learning process as it allows users to interact with the model and input data for prediction. Overall, these concepts are all interconnected, and their successful implementation leads to the creation of an accurate machine learning model capable of predicting outcomes based on the provided data. As shown in Figure 1

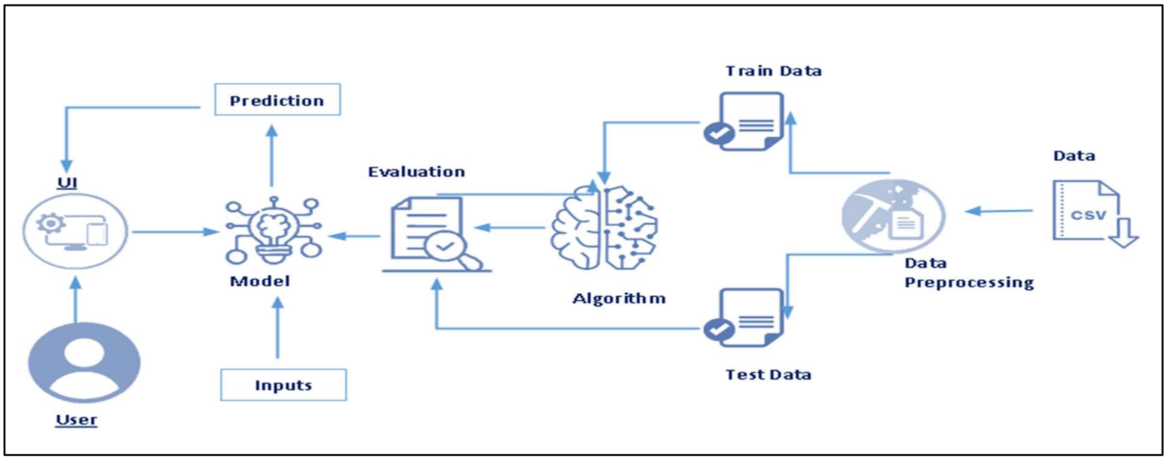


Figure 1 | Work Flow Diagram

Data Collection:

Age, gender, body mass index, blood pressure readings (both systolic and diastolic), heart rate, and serum glucose readings from glucometers are some of the variables that might help predict cardiac arrest. Wearable technology and health monitoring tools were used in this research to collect the aforementioned characteristics. The measuring device has the ability to record the vital signs needed to forecast cardiac arrest. the gathering of data in real time from different portable devices, including smart watches, sphygmomanometers, and glucose metres. Smart watches can be used to measure BMI and pulse rate. Sphygmomanometers are devices that measure both systolic and diastolic blood pressure. A glucometer may be used to check your blood sugar levels.

Dataset Description:

To be able to reach this goal, for each person there are fifteen attributes divided into three groups: demographic, medical history of the patient, and his or her current medical status (such as BMI, heart rate, etc.). Each attribute is a potential risk factor to determine if the patient will suffer a CHD in ten years. On the Kaggle website, the dataset for cardiac arrest is accessible to everyone. It contains the data of over 4,040 inhabitants of the Massachusetts town of Framingham. This dataset consist of 4,040 rows and 16 columns.

Here is the list of the features:

1. Category

- Gender: male (value 1) or female (value 0)
- age: Number of years as an integer number.

- Education: Some high school (value 1); some high school degree or GED (value 2); some college or vocational school (value 3); some college degree or vocational school degree (value 4).
- smoker: smoker (1) or nonsmoker (0)
- CigsPerDay: Average number of cigarettes smoked per day as an integer (0 for nonsmokers).

2. Medical history of the patient

- BPMeds: If the patient is on BP medication, represent 1, otherwise, 0.
- PrevalentStroke: if the patient already had at least one stroke, represent 1; otherwise, 0
- prevalentHyp: If the patient has hypertension, represent 1, otherwise 0.
- diabetes: if the patient has diabetes, represent 1; otherwise, 0.

3. current medical status of the patient

- totChol: float number of the cholesterol measured as mg/dL
- sysBP: float number of the systolic blood pressure measured in mmHg
- diaBP: float number of the diastolic blood pressure measured in mmHg
- BMI: body mass index (weight/height 22) of the patient measured as kg/mt 22. It is a float number.
- heartRate: Beats/Minute. It is an integer number.
- Glucose: Float the quantity of glucose in the patient's blood. Measured as mg/dL

As we said, this dataset is labelled. The column's name is TenYearCHD. There are two classes:

- Class 0: The patient won't have a CHD disease in 10 years.
- Class 1: The patient will have at least one CHD disease in 10 years.

Data preprocessing:

Preprocessing refers to the preparation of raw data before it is analyzed or used in a specific application. It involves a series of steps to clean, transform, and organise the data so that it can be more easily analysed or used in a machine learning algorithm. Preprocessing is an important step in data analysis as it can improve the accuracy of the results and make the data more meaningful and useful for the intended purpose. All entries in our dataset number 4240.

Importing all required libraries, loading the data from the csv file, and then displaying dataset information, which includes dataset entry numbers, datatypes, and columns, removing all duplicate rows from the dataframe after that, performing a statistical summary of the dataset, checking the missing values in the dataset, and checking the shape of the dataset after removing missing values Then, on the dataset attributes, perform an outlier detection. It is a fact or observation that significantly deviates from the established norm or average of the data collection. An outlier could result from pure chance, but it could also be a sign of measurement inaccuracy. Then, eliminate all outliers.

Algorithm Used:

Logistic Regression:

For binary and multiclass classification issues, logistic regression is a well-liked machine learning approach. It is a kind of linear model that predicts the likelihood of the target variable in light of the data provided as input. In order output predicate one or zero, logistic regression employs a logistic function. Using gradient descent optimisation methods and maximum likelihood estimation, the programme calculates the model parameters.[1]

Gradient Boosting Classifier

A machine learning technique called Gradient Boosting Classifier is a member of the family of gradient boosting methods that is employed for both classification and regression problems. The gradient-boosting classifier creates an ensemble of decision trees, each of which aims to fix the flaws in the one before it. The algorithm uses gradient descent optimisation to minimise the loss function and find the best fit for the training data.[6]

Random Forest Classifier (RFC)

It adheres to the idea of ensemble learning, which brings together various classifiers to address difficult issues and enhance model performance. It often entails a number of DTs constituting a subset of a particular dataset to increase the dataset's predictive accuracy. The Random Forest Classifier reduces the risk of overfitting and increases the model's accuracy and generalizability. The algorithm uses bootstrapping aggregation (bagging) to create diverse decision trees, which improves the model's performance.[1]

Decision Tree

A common machine the learning strategy for problems with classification and regression is the decision tree. Based on the input attributes, It creates a model of decisions and their results that resembles a tree. Considering the values of the characteristics that maximise the homogeneity of the target variable within the subsets, the algorithm divides the dataset into subsets. A decision tree can handle missing data as well as continuous and categorical input characteristics.[1]

Support Vector Machine

At its core, Finding a hypersurface that can precisely represent a hypersurface is the foundation of SVM. separate data points into different classes. The hyperplane is selected in a manner such that the margin between the two closest points from each class is maximized to a great extent. The data samples that are in close proximity to the hyperplane are referred to as support vectors, and they have a crucial impact on the algorithm's capability to extrapolate effectively to novel, unseen data.[6]

K-Neighbors Classifier

K-Neighbors Classifier (KNN) is an instance-based learning algorithm that utilises the entire training dataset to predict new data points based on the K closest neighbours, assigned the most frequent class label based on a similarity measure. To optimise KNN's performance, the GridSearchCV class performs a grid search over a specified range of hyperparameter values on a validation set. KNN can handle both categorical and continuous data and is easy to interpret, making it a popular choice for classification tasks.

Exploratory Data Analysis

Exploratory data analysis (EDA) is the process of examining and summarising a dataset's key features in order to understand the information and spot trends or linkages. By using this exploratory data analysis and applying it to our dataset, it shows in Figure 2. After that, we are using some statistical methods like Pearson, Kendall, and Spearman. Using this method, we are finding the correlation proportional to direct and inversly after that, removing the highly correlated same features in the dataset after finding the correlation. After that, plotting the correlation matrix and removing all dependent highly correlated values, it means the same values, then it is shown as Figure 3.

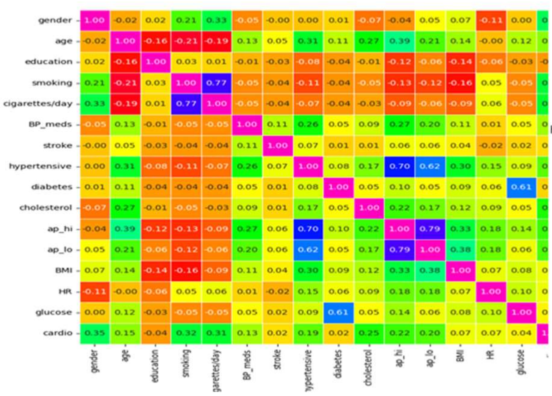


Figure 2 | Before correlation matrix.

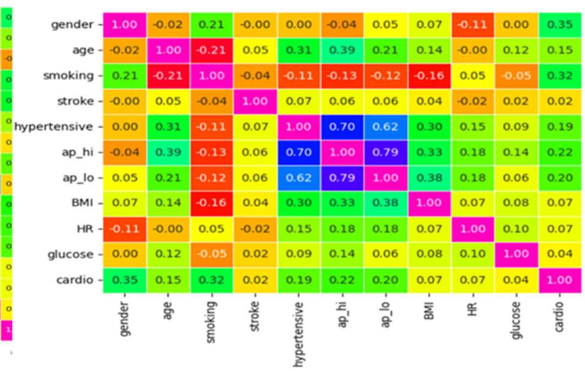


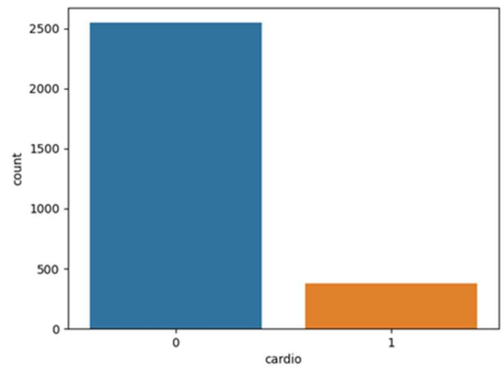
Figure 3 | After correlation matrix.

Scaling

Standard Scaler is a preprocessing technique that scales dataset features to have a standard normal distribution or a standard Gaussian distribution, making them comparable. It helps machine learning algorithms by preventing large-scale features from dominating the data. StandardScaler calculates the mean and standard deviation for each feature in the training data, and each value is transformed by removing the mean from it and dividing it by the standard deviation. The resulting feature has a standard normal distribution or a standard Gaussian distribution, simplifying data comparison and analysis. Scaling is an essential step in data preprocessing that enhances model performance.

Synthetic Minority Over-sampling Technique(SMOTE)

The SMOTE algorithm is used to address imbalanced datasets in machine learning. In a classification problem, An imbalanced dataset is one in which there are disproportionately fewer samples in one class than the other, leading to biassed models. The minority class is oversampled using SMOTE to provide synthetic instances. It uses an example from a minority class and locates its nearest neighbours. It then generates new examples by interpolating between the minority example and its k-nearest neighbours, creating synthetic examples similar to the minority class but with some differences. SMOTE increases dataset samples using oversampling to match the target and untargeted data. Cardio data feature analysis requires displaying datasets containing cardio events as counterplots before balancing the cardiac target dataset, as shown in Figure 4. Figure 5 shows the balanced cardiac target dataset after a counterplot using the SMOTE algorithm.



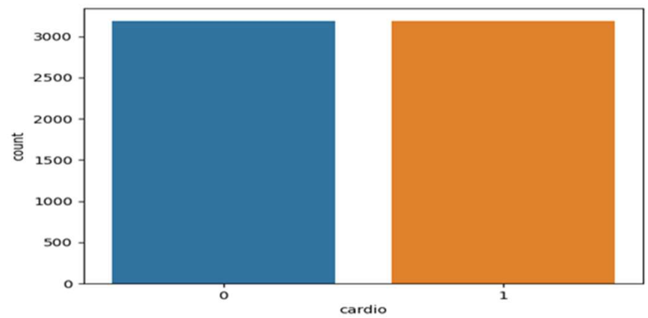


Figure 4 | Before over sampling cardio feature

Figure 5 | After oversampling cardio feature

4. Results and Analysis

The cardiac arrest dataset was analysed using various machine learning algorithms, and their performance was assessed using criteria including accuracy, sensitivity, precision, and F1-score. In this context, the numbers "0" and "1" denote the absence and presence of cardiac arrest, respectively. All metrics frequently used in data analysis and machine learning to assess a model's performance are presented in Figure 6. The above model evaluation the precision and f1-score are low values because of the numerator value less than the denominator value, while the model evaluation matrix shows on it.

Model	Accuracy	Precision	Recall	F1_score
Logistic Regression	0.840164	0.441489	0.873684	0.586572
Gradient Boosting Classifier	0.918033	0.625899	0.915789	0.743590
Random Forest Classifier	1.000000	1.000000	1.000000	1.000000
Decision Tree Classifier	1.000000	1.000000	1.000000	1.000000
Support Vector Machine	0.867486	0.494048	0.873684	0.631179
K-Nearest Neighbors	0.998634	1.000000	0.989474	0.994709

Figure 6 | Model Evaluation

The performance of several algorithms was evaluated using the confusion matrix, It includes elements with values like TP, TN, FP, and FN. These metrics were then used to calculate performance measures for the algorithms, and the results are shown in a visual format in Figure 7.

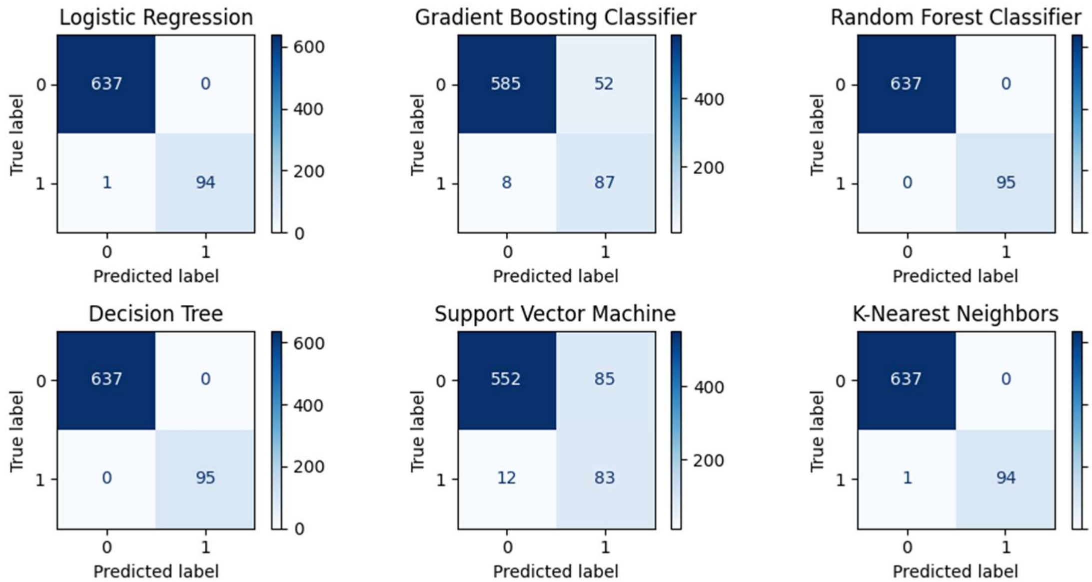


Figure 7 | Confusion matrix`s of different ML Algorithms.

A visual tool for evaluating a classification model's performance is the ROC (Receiver Operating Characteristic) curve. The model's efficacy can be seen visually by plotting the true positive rate against the false positive rate at various categorization levels. The model's overall performance over a variety of classification criteria is represented by the AUC curve of the ROC curve. A model is said to be flawless if it predicts all samples correctly with an AUC score of 1, while a score of 0 indicates a model that makes incorrect predictions for all samples. In our analysis, all classification models except for two achieved an AUC score of 1, indicating excellent predictive accuracy. Figure 8 illustrates the ROC curves for various machine learning algorithms, and Figure 9 shows all ROC curves plotted on a single graph.

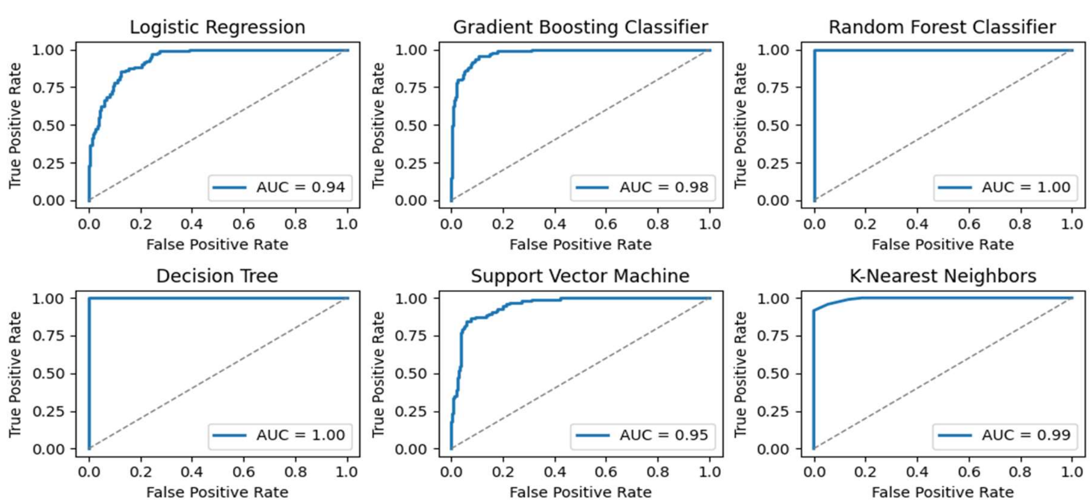


Figure 8 | ROC curve different ML Algorithms.

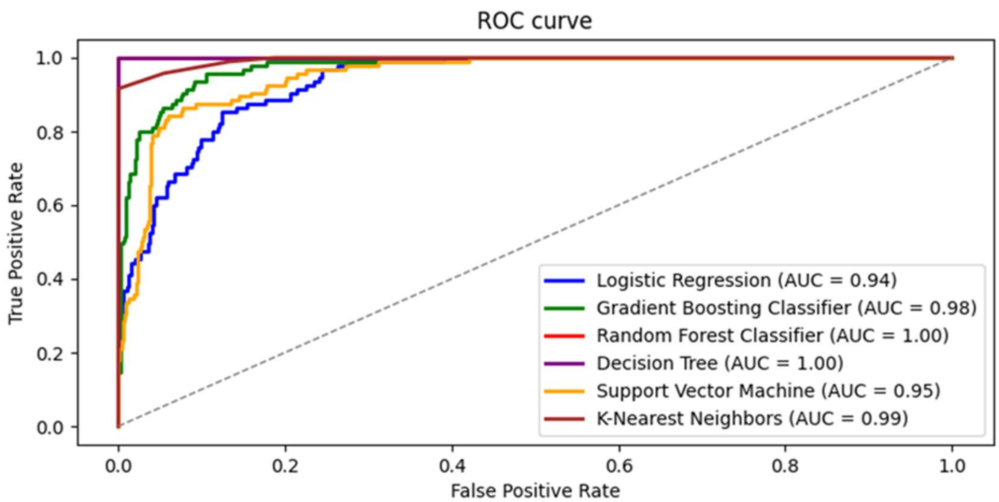


Figure 9 | ROC curve different ML Algorithms in a single graph.

5. Conclusion:

This work has improved the classification methods for detecting cardiac arrest. Our primary goal is to diagnose cardiac arrest in patients as quickly as possible using a machine learning model. We achieved this by collecting real-time data through sensors, displaying the data in an interactive graphical user interface, and providing a diagnosis through the application. In addition, other diseases can be incorporated into this study, and predictions can be made using various classification models. We can also increase the number of features and use a larger dataset to generate more accurate predictions.

References

1. Minsu Chae,Hyo-Wook Gil,Nam-Jun Cho and Hwamin Lee *Machine Learning-Based Cardiac Arrest Prediction for Early Warning System*. Mathematics 2022, 10, 2049.doi.org/10.3390/math10122049.
2. Joon-myoung,Youngam Lee,Yeha Lee,Seungwoo Lee,Jinisk Park *An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest*. J Am Heart Assoc. 2018;7:e008678. DOI: 10.1161/JAHA.118.008678.
3. Erik Alonso,Unai Irusta,Elisabete Aramendi and Mohamund R.Daya *A Machine Learning Framework for Pulse Detection Out-of-Hospital Cardiac Arrest*. IEEE Access. (2020) 10.1109/ACCESS.2020.3021310.
4. Joon-myoung Kwon,Kyung-Hee Kim,Ki-Hyun jeon,Soo Youn Lee Jinisk Park and Byung-Hee Oh *Artifical intelligence algorithm for predicting cardiac arrest using electrocardiography*. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine (2020) 28:98 doi.org/10.1186/s13049-020-00791-0.
5. L.Murukesh,M.Murugappan,M.Iqbal and Dr.K.Saravanan *Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features*. Journal of Medical Imaging and Health Informatics · August 2014 DOI: 10.1166/jmihi.2014.1287.

6. Shah A, Ahirrao S, Pandya S, Kotecha K and Rathod S *Smart Cardiac Framework for an Early Detection of Cardiac Arrest Condition and Risk*. Front Public Health (2021) 9:762303. doi: 10.3389/fpubh.2021.762303.
7. Kang J, Kim Y, Shin Y, Huh J, Hong S, Kim W. *Association Between Time to Defibrillation and Neurologic Outcome in Patients with In-Hospital Cardiac Arrest*. Am. J. Med. Sci.2019,8, 143-148
8. Florence D, Wulfran B, Alain C. *Cardiac arrest: Prediction models in the Early Phase ofhospitalization.Curr.Opin.Crit. Care* June 2019 DOI:10.1097/MCC.00000000000613 .
9. Shahidah N, Zhang T, Lin Z. *Prediction of Cardiac arrest in critically ill patients presenting to the Emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score* 2012 June. DOI: 10.1186/cc11396.
10. Chae M, Han S, Gil H, Cho N, Lee H. *Prediction of In-Hospital Cardiac Arrest Using Shallow and Deep learning* July 2021. DOI: 10.3390/diagnostics11071255.
11. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. *In-Hospital Cardiac Arrest*.March 2019. DOI: 10.1001/jama.2019.1696.
12. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP.*Derivation of a Cardiac arrest Prediction Model using ward vital signs.* June 2013 DOI: 10.1378/chest.12-1605.
13. Nadkarni VM, *First documented rhythm and clinical outcome from in-hospital Cardiac arrest among children and adults*. Jan 2006 DOI: 10.1001/jama.295.1.50.
14. Sarmah SS. *An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network*. IEEE Access. (2020) 8:135784–97. doi: 10.1109/ACCESS.2020.3007561.
15. https://www.heart.org/en/health-topics/cardiac-arrest.
16. https://en.wikipedia.org/wiki/Cardiac_arrest#See_also.
17. https://www.nhlbi.nih.gov/health/cardiac-arrest.