PREDICTION OF CARDIAC ARREST USING MACHINE-LEARNING ALGORITHMS

Ananta Verma[1], Gaddam Sesha Sai Srikar[2], Hrushikesh Andurlekar[3] and Kapileswarapu Chandra[4]

[1]Department of CSE & Final year student of B-Tech Mechanical, Symbiosis Institute of Technology, Pune, India.
[2]Department of CSE & Final year student of B-Tech Mechanical, Symbiosis Institute of Technology, Pune, India.
[3]Department of CSE & Final year student of B-Tech Mechanical, Symbiosis Institute of Technology, Pune, India.
[4]Department of CSE & Final year student of B-Tech Mechanical, Symbiosis Institute of Technology, Pune, India.

Abstract

Cardiac arrest is a cardiovascular disruption resulting from an electrical current that stops the heart from beating. For effective cardiac arrest resuscitation, accurate identification of high-risk patients, proper preparation, and rapid commencement of clinical care are critical components. Among these stages, precisely identifying patients at high risk of cardiac arrest is a primary technique, and numerous research studies have been undertaken to predict the risk of cardiac arrest. Machine learning starts with individual observations, then automatically searches multivariate data, derives trustworthy outcome predictions, and finally builds reliable models. Machine learning is widely regarded as an essential tool for dealing with complicated problems in science, particularly in biomedical and astronomical research.

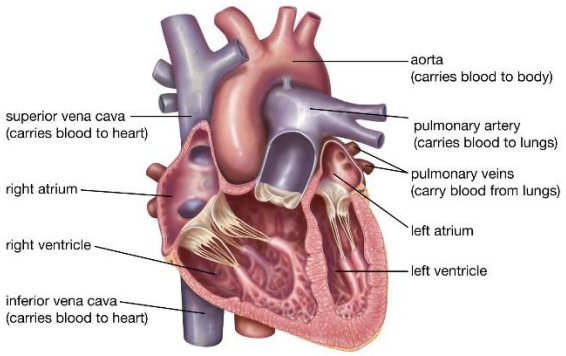
For this paper we will be using 4 different ML algorithms namely Random Forest, KNN, XGBoost and Logistic Regression. Regardless of the potential benefits of machine learning algorithms, various considerations are to be considered while developing a viable algorithm for forecasting cardiac arrest. In this paper we will further explore the potential of the aforementioned algorithms.

Introduction

During earlier times for the prediction of cardiac arrest Modified Early Warning Score (MEWS) was used, MEWS can be performed on all hospitalized patients to detect clinical deterioration and the possible requirement for a higher level of care.

A cardiac arrest is a cardiac disruption caused by an electrical current that causes the heart to cease beating. Accurate identification of high-risk patients, thorough planning, and fast initiation of clinical care are important components for efficient cardiac arrest resuscitation. Among these stages, a fundamental strategy is precisely identifying patients at high risk of cardiac arrest, and several studies have been conducted to forecast the risk of cardiac arrest. Machine learning begins with individual observations, then explores multivariate data automatically,

generates reliable outcome predictions, and finally builds trustworthy models. Machine learning is widely regarded as a crucial technique for dealing with complex scientific problems, notably in biomedical and astronomical research. The overwhelming number of OHCAs take place in people's homes or apartments (68.5%), followed by public settings (21%), and nursing facilities (10.5%). A bystander witnessed cardiac arrest in 37% of cases and an EMS provider in 12% of cases.



Other than this there are Rapid Response Team (RRT) activation criteria and Cardiac Arrest Risk Triage (CART) scores to predict cardiac arrest was developed. These were better than the MEWS in terms of accurate prediction. In recent times ML algorithms and deep learning models are used for this purpose. These ML and DL algorithms have better accuracy and show a better confidence ratio than the earlier models.

The subject is still under study and various new discoveries are made with the advancement of monitoring technology and newer AIML models are giving better results.

Methodology

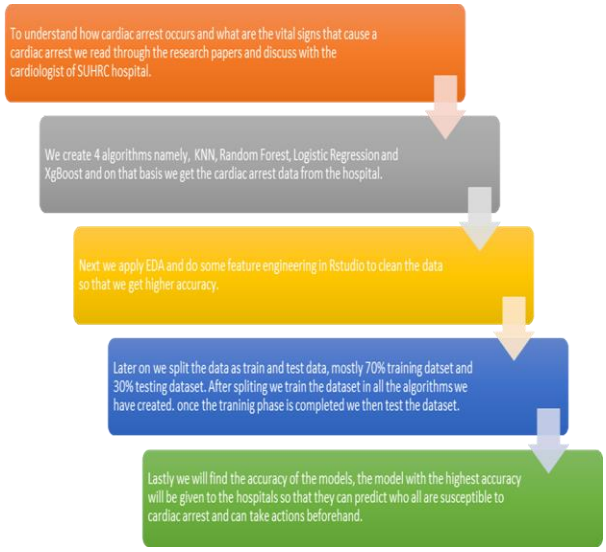


Figure. 1

1. Random Forest

Random Forest belongs to a supervised machine learning method which makes it a popular ML algorithm which is based on the concept of ensemble learning. This process is a combination of multiple classifiers to solve a complex problem and to improve the performance of the model. Greater number of trees in the forest, higher the accuracy and prevention of overfitting.

Working of Random Forest:

a. Picking any random K data points from the training set

b. Create the decision trees linked to the chosen data points (Subsets).

c. Pick the decision tree N that you intend to construct.

d. Repeat a & b

e. When dealing with fresh data points, locate each decision tree's predictions and categorize them according to whatever group receives the most votes.

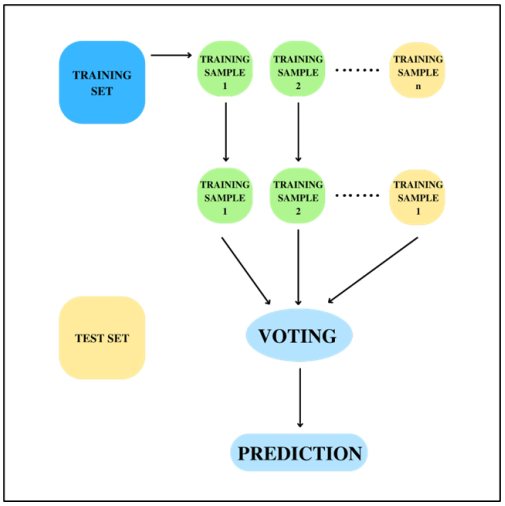


Figure. 2

Mathematical Formula

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

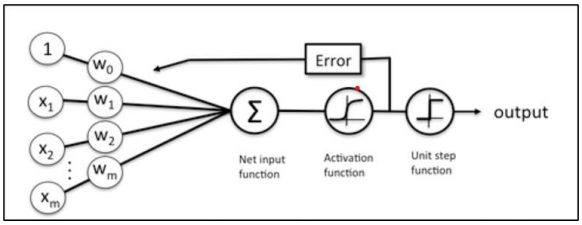
Calculation of each feature in a decision tree is calculated as

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k}$$

2. Logistic Regression

Logistic Regression has the ability to provide probabilities and classify new data using continuous and discrete datasets. Using a predetermined set of independent factors, it is used to predict the categorical dependent variable. Since it anticipates the output of a categorical dependent variable, its result must be categorical, i.e., either True or False, 0 or 1.
For logistic regression to be predicted, we need:
A large number of data sets.
A dependent variable which should be in binary.
The variables should be understood.

As mentioned above the binary regression the dependent variable (1) should give the desired output.



$$\mathbf{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

It is shown that y is the predictable variable.

And beta is the predictors coefficient given x is the independent variable.

$$p = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)}}{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)}+1}$$

or

$$p = \frac{e^{Y}}{e^{Y}+1}$$

where p is the probability of success. With little further manipulations, we hav

$$\frac{p}{1-p} = e^{Y}$$

and

$$\log \frac{p}{1-p} = Y$$

If we remember what was $\mathbf{Y}$, we get

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

3. KNN

K- Nearest Neighbors (KNN) algorithm is a supervised machine-learning algorithm, which is also known as a lazy algorithm as it does not learn much from its training data. It is mostly used for classification data but can also be used for regression.
Steps to create the algorithm:

- Find the value of K. To do so, we take the square root of the number of rows and always take the value as an odd number so that the algorithm does not get

confused between the even number of classes.

- If we take the value of k to be very low or 1 then the model turns out to be underfit. Similarly, if we take a very large value of k, then the model will turn out to be overfit and expensive, as it will be calculating all of the relationships inside the dataset.
- Decide which distance metric or algorithm to be used. Here we go for the Euclidean distance as it is the most popular and gives the highest accuracy. We can use other distances: Manhattan distance, Minkowski distance and cosine similarity etc.

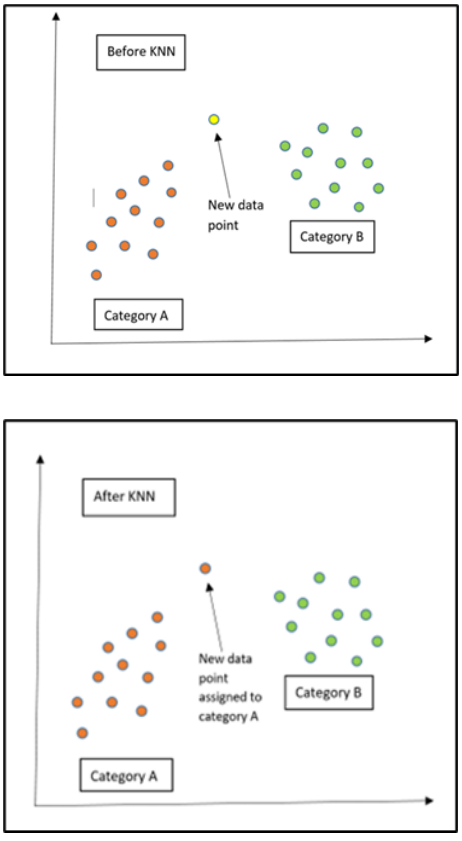


Figure. 3

Euclidean Distance is a very common and popular distance measure but it is limited to real value vectors.

Euclidean formula: d (x, y) = d (x, y) =

$\sqrt{\sum} (yi - xi)^2$

Manhattan distance is a distance metric that measures the absolute values between two points.

Manhattan Distance: d (x, y) = d (x, y) =

$\sum |yi - xi|$

Minkowski distance is the generalized form of the above two distances. The value of p here in Euclidean distance becomes two whereas in Manhattan distance becomes one.

Minkowski distance: $\sum |yi - xi|^{1/p}$

4. XG Boost

XGBoost is a distributed gradient boosting library that has been developed to be very effective, adaptable, and portable. It uses the Gradient Boosting framework to construct machine learning algorithms. A parallel tree boosting method called XGBoost (also known as GBDT or GBM) is available to quickly and accurately address a variety of data science issues. The same algorithm can answer problems with more than a trillion instances and runs on key distributed environments (Hadoop, SGE, MPI). Extreme gradient boosting uses a technique called batch gradient descent to train many models simultaneously. Each model in the

ensemble is a prediction of the class label for the input sample. The models are trained on the training dataset and evaluated on their test dataset. Each model is trained using a different optimization algorithm and has a slightly different architecture

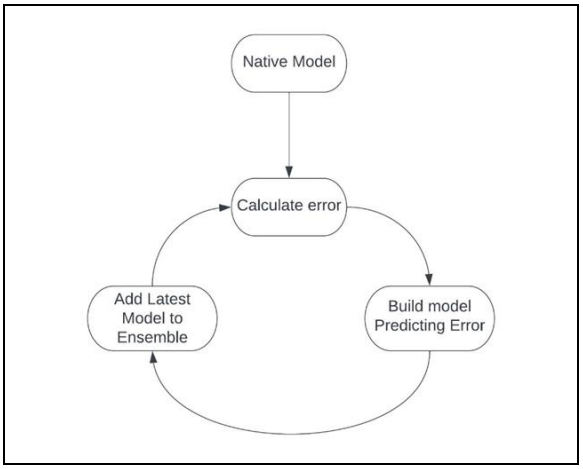


Figure. 4

5. Linear Regression

It is a statistical technique for performing predictive analysis. For real, numerical, or continuous data, linear regression generates predictions. The linear regression procedure, often known as linear regression, demonstrates a linear relationship between a dependent (y) and one or more independent (y) variables. Given that linear regression demonstrates a linear connection, it may be used to determine how the dependent variable's value changes as a function of the independent variable's value.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots$$

y is a dependent variable
x1, x2, are explanatory variables
b1, b2 explains the correlation of explanatory variables with dependent variable.

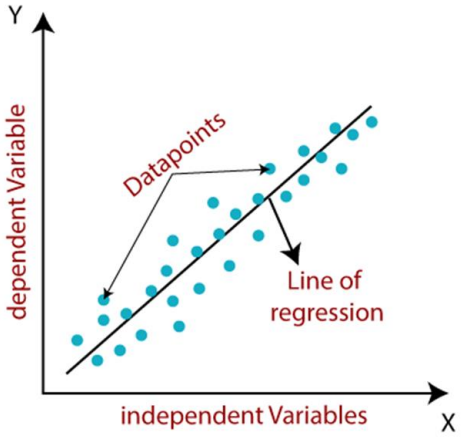


Figure. 5

6. Naïve Bayes

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

Working of Naïve Bayes: -

1. Convert the given dataset into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

As this algorithm is related to Bayes Theorem the formula used is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Future Scope

- Working together with the cardiology department to create an extensive database for research.
- Research on different algorithms to figure out which algorithm is the most suitable for this application.
- Create an application that can be used for data collection and testing, using the said algorithm.

Conclusion

From the model with highest accuracy, we will predict the risk of cardiac arrest of the patient. This will help the doctors and hospital to understand the patient's situation and help the doctors administer precautionary measures for the same.

Related Work

S.NO	NAME OF PAPER	DATA SAMPLE	ALGORITHM USED	ACCURACY
1.	Artificial Intelligence Algorithm for predicting cardiac arrest using ECG	47,505 ECGs of 25,672 adult patients admitted to two hospitals, who underwent at least one ECG from October 2016 to September 2019.	AUROC(De-Long's method) AUROC: Area under the receiver operating characteristics curve	AUROC-0.91 – 0.94
2.	Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes	166 Acute Coronary Syndrome patients who had in-hospital cardiac arrest.	The XGBoost model. The K-nearest neighbor model	AUC (0.958 [95%CI: 0.938-0.978]), accuracy (88.9%), sensitivity (73%), negative predictive value (89%), and F1 score (80%) The K-nearest neighbor model specificity (99.3%) and positive predictive value (93.8%)

3.	Cardiac Arrest Prediction using Machine Learning Algorithms	a prospective and population-based, nationwide registry of Oklahoma HealthCare Authority in Japan and a weather forecast dataset from a Weather Company in United States.	Logistic Regression Decision Tree Random Forest Support Vector Machine	- 57.31 % - 50.5% - 60.5% - 53.8%
4.	Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features.	MIT/BIH SCD database of 23 subjects and 2 leads of 250 Hz sampling frequency and 20 people had SCA, 3 was paced-13 Males and 8 Females and 2 Unknown. MIT/BIH NSR database of 18 subjects and 2 leads of 128 Hz sampling frequency and all in sinus rhythm, -5 Males and 13 Females (Holter recording in both)	Sequential Feature Selection algorithm is used to identify the ideal values for prediction using TreeBagger Classifier. and One-way ANOVA. SVM (Support Vector Machine)	- 96.36%

	An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest	This retrospective cohort study of patients who were admitted to 2 hospitals from June 2010 to July 2017 of total 52 131 patients	Logistic Regression Random Forest	- 40% - 80%

Proposed Work

Algorithms	Description	Accuracy in papers
KNN	KNN is a non-parametric algorithm as it doesn't rely on a set of weights and biases and has any assumptions about the data. KNN is commonly used for its ease of interpretation and low calculation time. K is the number of data points considered near the observation. The distance between the new data point and the nearest neighbors is calculated using Euclidean distance.	86.2%
Random Forest	Random Forest has the ability to perform both regression as well as classification which is done using a technique known as Bootstrap and Aggregation along with the use of multiple decision trees. Bootstrap is basically in the case of classification where we get the final output from the method of majority voting classifier. As the name states, Random Forest consists of a large number of trees. Each tree consists of a large number of branches which is considered a class prediction.	82.6%
Logistic Regression	It is the machine learning algorithm used to predict the outcome using 0's and 1's (dependent variable). It is a classification algorithm used to assign observations to a discrete set of classes. transforms its output using the logistic or sigmoid function to	84%

	return a probability value which can then be mapped to two or more discrete classes.	
XG Boost	Extreme gradient boosting is a simple and powerful algorithm for learning multiple output regression and classification models. XG boost is based on this model. Extreme gradient boosting uses a technique called batch gradient descent to train many models simultaneously. Each model in the ensemble is a prediction of the class label for the input sample. The models are trained on the training dataset and evaluated on their test dataset.	88.8%

References

[1]https://www.heart.org/en/health-topics/heart-attack/about-heart-attacks/heart-attack-or-sudden-cardiac-arrest-how-are-they-different
[2]https://www.healthline.com/health/cardiac-arrest#diagnosis
[3]https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983
[4]https://www.nhs.uk/conditions/electrocardiogram/
[5]https://www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668
[6]http://mayoclinic.org/tests-procedures/holter-monitor/about/pac-20385039
[7]https://www.geeksforgeeks.org/random-forest-regression-in-python/
[8]https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/
[9]https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark
[10]https://www.javatpoint.com/machine-learning-random-forest-algorithm
[11]https://medium.com/analytics-vidhya/evaluating-a-random-forest-model-9d165595ad56
[12]https://mljar.com/blog/feature-importance-in-random-forest/
[13]https://www.youtube.com/watch?v=_cDl0TFICJY
[14]https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning
[15]https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.
[16]https://youtu.be/4HKqjENq9OU
[17]https://www.ibm.com/in-en/topics/knn
[18]https://www.javatpoint.com/linear-regression-in-machine-learning
[19]https://www.javatpoint.com/machine-learning-naive-bayes-classifier
[20] Andersen, L. W., Kim, W. Y., Chase, M., Mortensen, S. J., Moskowitz, A., Novack, V., ... & Donnino, M. W. (2016). The prevalence and significance of abnormal vital signs prior to in-hospital cardiac arrest. Resuscitation, 98, 112-117.
[21] Chae, M., Han, S., Gil, H., Cho, N., & Lee, H. (2021). Prediction of in-hospital cardiac arrest using shallow and deep learning. Diagnostics, 11(7), 1255.

[22] Fieselmann, J. F., Hendryx, M. S., Helms, C. M., & Wakefield, D. S. (1993). Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. Journal of general internal medicine, 8(7), 354-360.
[23] Karthikeyan, R., Babu, D. V., Suresh, R., Nalathambi, M., & Dinakaran, S. (2021, July). Cardiac Arrest Prediction using Machine Learning Algorithms. In Journal of Physics: Conference Series (Vol. 1964, No. 6, p. 062076). IOP Publishing.
[24] Kim, J., Chae, M., Chang, H. J., Kim, Y. A., & Park, E. (2019). Predicting cardiac arrest and respiratory failure using feasible artificial intelligence with simple trajectories of patient data. Journal of clinical medicine, 8(9), 1336.
[25] Kwon, J. M., Kim, K. H., Jeon, K. H., Lee, S. Y., Park, J., & Oh, B. H. (2020). Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. Scandinavian journal of trauma, resuscitation and emergency medicine, 28(1), 1-10.
[26] Kwon, J. M., Lee, Y., Lee, Y., Lee, S., & Park, J. (2018). An algorithm based on deep learning for predicting in-hospital cardiac arrest. Journal of the American Heart Association, 7(13), e008678.
[27] Murukesan, L., Murugappan, M., Iqbal, M., & Saravanan, K. (2014). Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features. Journal of Medical Imaging and Health Informatics, 4(4), 521–532.
[28] Oh, H., Lee, K., & Seo, W. (2016). Temporal patterns of change in vital signs and cardiac arrest risk triage scores over the 48 hours preceding fatal in-hospital cardiac arrest. Journal of advanced nursing, 72(5), 1122-1133.
[29] Popescu, D. M., Shade, J. K., Lai, C., Aronis, K. N., Ouyang, D., Moorthy, M. V., ... & Trayanova, N. A. (2022). Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nature Cardiovascular Research, 1(4), 334-343.
[30] Wu, T. T., Lin, X. Q., Mu, Y., Li, H., & Guo, Y. S. (2021). Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. Clinical cardiology, 44(3), 349-356