International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 03 | March - 2022 ISSN: 2582-3930

Analysis of Heart attack Prediction Using Supervised Machine Learning Algorithms with an Electrotherapeutic Approaches

Sasikala P [1], Venkatesh R[2]

P.G Student, Department of Computer Science and Engineering,

Sri Shakthi Institute of Engineering and Technology, Coimbatore, India[1]

Assistant Professor, Department of Computer Science and Engineering,

Sri Shakthi Institute of Engineering and Technology, Coimbatore, India[2]

Abstract -In recent days, heart attacks are considered to be the biggest concern in the health care industry. In most of the times, it is quite difficult for doctors to identify the disease. So, it results in more death of people. If the disease can be identified at an earlier stage, it will be easier to prevent the death rate. It is impossible for an individual to frequently undergo costly tests like ECG. Prediction of heart attack requires more precision, perfection and accuracy. A little mistake can cause unwanted problem or death of the person, there are numerous death cases related to heart which are increasing gradually day by day. As the medical diagnosing is a decision-making technique, an intelligent decision system can be implemented by using various machine learning techniques. Machine Learning (ML), a part of Artificial Intelligence technique (AI) which provides a support in predicting data. The use of machine learning technique is gradually increasing in the medical field. The main aim of this paper is to reduce the efforts and risk made by doctors in predicting heart attacks. One of the biggest challenges in healthcare is to record and analyse huge amount of information about patients. The main objective is to investigate different factors and its impacts in identifying heart attack. The factors such as chest pain, age, gender, fasting blood sugar level and cholesterol level. In this paper an analysis is done to check the performance of various heart disease prediction techniques, namely k-nearest neighbour (KNN), random forest,decision tree, linear regression and support vector machine (SVM).In this research, an accuracy of various machine learning algorithms for predicting heart attacks are obtained.

Key Words: Machine learning (ML), Artificial Intelligence (AI), K-Nearest Neighbour (KNN), Support Vector Machine (SVM).

1.INTRODUCTION

Heart diseases are considered to be one of the biggest concerns in the medical field with its number of deaths. As per the report of WHO, heart disease is considered to be the major asymptomatic diseases. Almost 10% of the world are getting trouble due to this heart disease. In the world around 12 million deaths are happened every year. Irregularity in the blood flow in and out of the heart causes heart attack. The prediction made by doctors is not always100% accurate. Accurate Identification of heart disease became very crucial nowadays. Manual observation is not always accurate for identification of disease. Hence there is a need to transform clinical diagnosis into computational diagnosis. This computational diagnosis can be used to find the crucial attributes of heart disease. Lots of efforts has already been done to predict the heart disease using the ML algorithms. Today, most patients die because their diseases are discovered at an advanced stage due to instrument inaccuracy, hence there is a need to learn about more effective algorithms for disease prediction. Due to advancement in technology, humans are under safe zone so that most of the works are being done by machines. If the disease can be identified at a premature stage perfectly followed by appropriate treatment then a number of lives can be saved. The availability of huge amount of data in the medical field really helped researchers to detect useful solution for complex disease identifications. Medical diagnosis is mainly coming under classification-based problem. With the advancement in machine learning, several classifiers and clustering algorithms are being used. Supervised, Unsupervised Machine Learning, and Reinforcement Learning are the three types of machine learning. Under machine learning so many approaches like Deep learning, Reinforcement Learning and Evolutionary learning which are nowadays becoming very much popular for its fast detection and accurate prediction. As a result, we propose gathering relevant data on all aspects of our field of research, training the data using the proposed machine learning method, and predicting the likelihood of a patient contracting a cardiac ailment.

The main aim of this paper is

a) To analyse and understand the work done by various researchers for the classification of heart disease.

b) By taking some vital factors to analyse the accuracy of all the machine learning algorithms.

2. RELATED WORK

1] A state of Art analysis in “Heart Disease Prediction using Machine Learning Techniques” by Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta. It gives an idea in using various machine learning algorithms to predict heart attack. It also helps in comparing accuracy among the algorithms. It supports the doctors to identify the diseases at an earlier stage and helps in saving many human lives.

2] "An Investigation of Heart Disease Prediction Techniques" by Shanmugasundaram G, Malar Selvam V, Balaji. It helps in prediction of heart attack based on certain factors. The factors are classified into common factors and medical oriented factors. With the knowledge of factors listed prediction can be done easily.

3]" Heart Attack Prediction System" by Sushmita Manikandan. It helps in identifying the deadliest disease at an earlier stage. It is used for handling a large volume of data. It uses a supervised machine learning algorithm to calculate the accuracy.

4] "A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning" by Debabrata Swain, Santosh Kumar Pani, Debabala Swain. It supports the use of dimensionality reduction techniques to predict diseases and calculate accuracy. It also helps in using comparison factor like number of output classes which gives best accurate results.

5] "Application of Machine Learning in Disease Prediction" by Pahulpreet Singh Koh, Shriya Arora. It gives an idea to use machine learning algorithms in healthcare. It helps in disease identification and diagnosing. It supports users to apply machine learning techniques in medical sector in risk predictions. It predicts the importance of machine learning methods in healthcare industry.

3. PROPOSED MECHANISMS

1.MACHINE LEARNING TECHNIQUES

Machine learning is an efficient technology which involves ability of a machine to imitate human behaviour. Machine Learning mainly involves two main processes testing and training. The training is done with the help of available data and are tested based on the trained dataset. Machine Learning Algorithms are classified into three main categories namely: Supervised learning, Unsupervised learning, Reinforcement learning.

Parameter	Supervised learning	Unsupervised learning
Dataset	Labelled	Unlabelled
Method of Learning	Learning that is guided	Using a dataset, the algorithm learns on its own.
Complexity	Methods with less complexity	Complex in terms of computation
Accuracy	Higher Accuracy	Lower Accuracy

Table 1: shows some significant differences between supervised and unsupervised learning.

In this paper, an implementation of supervised learning algorithms are made as it provides more accurate results compared to unsupervised learning algorithms.

1.1 SUPERVISED MACHINE LEARNING ALGORITHMS

Supervised learning is defined as a learning under a proper guidance. Supervised learning involves a proper guide to train the dataset. Supervised Learning is when you consider that your learning is being led by a teacher. We have a dataset that serves as a teacher, and its job is to train the machine or model. When additional data is supplied to the model after it has been trained, it can begin generating predictions or decisions.

Supervised learning is further divided into the categories listed below.

i.Decision Tree
ii.Random Forest
iii.Support Vector Machine
iv.Linear Regression
v.K-Nearest Neighbor

1.2.SUPPORT VECTOR MACHINE

Support Vector Machine is a machine learning classification approach for analysing data and discovering patterns in classification and regression analysis. When data is classified as a two-class issue, SVM is usually considered. Data is defined in this method by determining the optimal hyper plane that isolates all data points from one class from the other. The greater the separation or edge between the two classes, the better the model is thought to be. Support vectors are data points that are located at the margin's edge. The mathematical methods used to design complex real-world problems form the foundation of SVM. Other approaches accessible in SVM include quadratic programming, sequential minimum optimization, and least squares, in addition to the kernel's capabilities. The most difficult part of constructing a model with SVM is avoiding overfitting and underfitting by choosing the right kernel and technique. Because our dataset has a large number of parameters and instances. As a result, we had the option of using the RBF or linear kernel. As a result, SVM's final model must be verified and validated against real-world data.

1.3 DECISION TREE

The Classification models are created using the Decision Tree technique in Machine Learning. The tree-like structure underpins this categorization approach. This is classified as supervised learning since the desired outcome is already known. The Decision tree technique may be used with both category and numerical data. The root node, branches, and leaf nodes make up a decision tree. The traversal path from the root to a leaf node is used to assess the data. They may have come to a good or negative conclusion on the risk of heart disease. These were compared to the real parameters to see if there were any false positives or negatives, demonstrating the model's accuracy, specificity, and sensitivity.

1.4 RANDOM FOREST

Random forest is one of the supervised machine learning algorithms, which is more flexible and produces a result with high accuracy. It is one of the most preferred machine learning algorithms in almost all the fields including healthcare. Random forest is similar to decision tree approach where it combines multiple decision trees to get accurate results. The main advantages of random forest are that can be applied to classification and regression approaches. An

important aspect of random forest is ability to understand the importance of features used. Overfitting problem is rare in case of random forest algorithm if there are enough trees in the forest. Even though the random forest algorithms provide accurate results, the large number of trees may result in slower computation of an algorithm. In prediction of heat attack random forest provides a highest accuracy compared to all other algorithms.

1.5 LINEAR REGRESSION

Linear Regression is a supervised machine learning algorithm. It is one of the simplest approaches which establishes a relationship between a dependent variable and an independent variable. It is used in the predictive analysis for making predictions for continuous variables namely age, salary, etc. It establishes a linear relationship by defining how the impact of dependent variable gets changed if the independent variables undergo any changes.

1.6 K-NEAREST NEIGHBOUR

The K-Nearest Neighbor is a supervised machine learning algorithm. It works on the basis of distance between data locations, and separate data's that are categorized based on this distance. All other groups of data are referred to as neighbors, and the number of neighbors is determined by the user. The k-Nearest Neighbor algorithm compares an unknown data to the k training data's that are the unknown data's closest neighbors. Finding the k nearest training Examples is the first step in applying the k-Nearest Neighbor algorithm to a new dataset. The distance between the unknown data and the training data can be calculated using many metrics, such as the Euclidean distance. Because distances are frequently dependent on absolute values, normalizing data before training and implementing the k-Nearest Neighbor method is advised. In the second phase, the k-Nearest Neighbor method uses a majority vote among the detected neighbors to classify the unknown data. It may be beneficial to weight the contributions of neighbors so that the neighbors who are closer give more to the average than those who are farther away.

4.METHODOLOGIES

Processing of system start with the data collection for this we use the Kaggle dataset which is well verified by number of researchers.

A. *DATA COLLECTION*

The Heart Disease Dataset from Kaggle has been used in our research work. The dataset provides the patients'information that consists of 303 records with 14 attributes. Each attribute is a potential risk factor and the descriptions of them are shown in Table.

B. ATTRIBUTE SELECTION

Attribute of dataset are property of dataset which are used for system and for heart many attributes are like heart bit rate of person, gender of the person, age of the person and many more shown in table for predication system.

SI. NO	**ATTRIBUTE SELECTION**	**DESCRIPTION**
1	Age	Patient's age (20 to 80)
2	Sex	GENDER OF PATIENT (male-0 female-1)
3	Cp	CHEST PAIN TYPE 0 represents typical angina 1 represents atypical angina 2 represents non-anginal pain 3 represents asymptomatic
4	Trestbps	RESTING BLOOD PRESSURE (in mm Hg on admission to hospital ,values from 94 to 200)
5	Chol	SERUM CHOLESTEROL IN MG/DL, values from 126 to 564)
6	Fbs	FASTING BLOOD SUGAR>120 MG/DL, true-1 false-0)
7	RestingECG	Resting electrocardiographics result (0 to 1) 0 represents normal 1 represents having ST-T wave abnormality 2 represents Possible or definite left ventricular hypertrophy
8	ThaliACH	Maximum heart rate achieved (71 to 202)
9	Exang	EXERCISE INCLUDED AGINA (1-yes 0-no)
10	Oldpeak	ST depression introduced by exercise relative to rest (0 to .2)

C. PREPROCESSING OF DATA

Pre-processing needed for achieving prestigious result from the machine learning algorithms. Data is often incomplete and inconsistent. Inconsistent and incomplete dataset significantly influence the performance of machine learning algorithms. The heart disease data set has been checked for missing values and null values as null value considerably effect on the conclusion that drawn from the data. There were no missing values in our dataset. As the percentage of missing value in our dataset is zero, we further moved to classification process. The below figure describes the methodologies to predict disease.

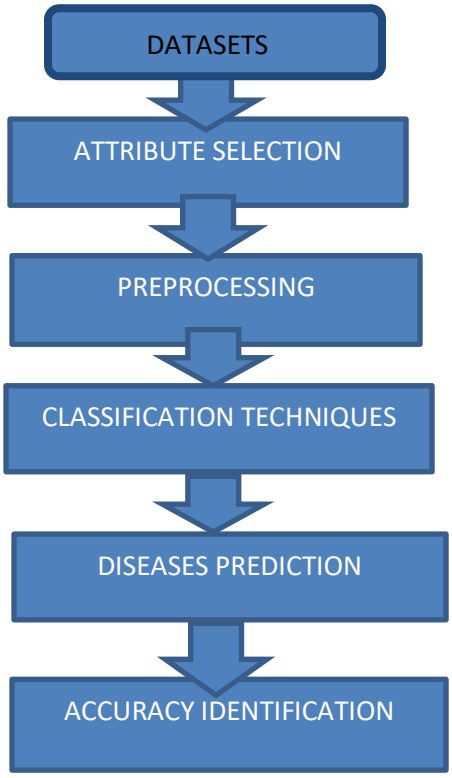

Fig 1. Shows the methodologies of heart diseases prediction

5.EXPERIMENTAL RESULTS

It has been observed that at the end of our experiment, result shows that Random Forest performed very well in comparison to SVM. linear regression, k-nearest neighbour and Decision tree.

ALGORITHMS	ACCURACY
SUPPORT VECTOR MACHINE	83%
DECISION TREE	79%
LINEAR REGRESSION	78%
K-NEAREST NEIGHBOUR	87%
RANDOM FOREST	95%

6.. DISCUSSIONS

- Few researchers have utilised big data sets for categorization, but they did not use adequate error management mechanisms. Error-handling methods such as removing the data row, using global values to fill in the empty values, and using factor mean to calculate the values may have been utilised.
- Some academics have rationalised their suggested algorithm as being more efficient than other algorithms without doing a thorough examination of alternative efficient algorithms.
- According to the findings of the inquiry, the researchers did not examine numerous factors for prediction. Many significant aspects might have been added, such as drinking and smoking. The majority of them have simply considered common characteristics.
- The attribute size is determined to be decreased in order to improve accuracy. This decrease in the number of characteristics does not properly indicate heart disease. All of the factors that influence cardiac disease should be taken into account.

7.CONCLUSION

A review of various machine learning algorithms for predicting cardiac disease is presented in this work. We also compared the categorization accuracy of all of the strategies presented.Major influencing elements for determining heart disease, as well as diverse research efforts in predicting heart disease, have been identified and documented in this survey. It has been found that not all attributes are considered by all researchers. Few attributes are deleted by a few studies in order to improve accuracy. The goal of this study is to compare the results of various supervised machine learning algorithms that can be used in heart attack prediction systems. Heart disease is predicted using decision trees, SVMs, KNNs, linear regression, and random forests in this study. Our report's main goal was to investigate and analyse the reasons behind different algorithms' variations. We took the Kaggle dataset for cardiac disorders, which has 303 occurrences, and divided the data into two halves, training and testing datasets, using a percent split. To test the accuracy, we looked at 14 distinct attributes and used five different methods. The dataset used for training and testing purposes determines the accuracy of machine learning algorithms. We had a thorough discussion on the major obstacles of various research projects for heart disease prediction that have yet to be addressed. In the future, researchers should incorporate all of the elements for identifying heart disease using an effective algorithm, as heart disease prediction is an essential concern for humans, and algorithm accuracy is one of the parameters for analyzing algorithm performance. In the future, more machine learning approaches will be used for the best analysis of heart diseases and for earlier disease prediction, so that the rate of death cases can be reduced through disease awareness. Following the completion of this comprehensive analysis, further work can be focused on implementing a competent technique that can predict heart disease with greater accuracy than all of the previously stated techniques. According to the findings of the experiments, the random forest classifier is more accurate than Decision Tree, SVM, Linear regression, and KNN.

REFERENCES

[1] Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta," Heart Disease Prediction using Machine Learning Techniques",
2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),2020.
[2] Santhana Krishnan J and Geetha S, "Prediction of Heart Disease using Machine Learning Algorithms" ICIICT, 2019.

[3] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2018.

[4] V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction," vol. 38, no. 3, pp. 124–128, 2016.

[5] P. Sujatha, K Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease", *IEEE International Conference for Innovation in Technology (INOCON),2020.*

[6] Sushmita Manikandan," Heart Attack Prediction System" International Conference on Energy, Communication, Data Analytics and Soft Computing,2017.

[7] M. Schaldach and H. Hutten, "New concepts in electrotherapy of the heart," *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1992, pp. 705-706, doi:10.1109/IEMBS.1992.5761180.

[8] Debabrata Swain, Santosh Kumar Pani, Debabala Swain," A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning", IEEE,2018.

[9] Pahulpreet Singh Koh, Shriya Arora," Application of Machine Learning in Disease Prediction", *4th International Conference on Computing Communication and Automation (ICCCA),2018.*

[10] Archana Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", International Conference on Electrical and Electronics Engineering,2020.

[11] Sharmila S et al., "Analysis of Heart Disease Prediction using Data Mining Techniques", International Journal of Advanced Networking & Applications (IJANA), 2017.
[12] D. Wilton, "How physiotherapy generates problems," *IEE Colloquium on Electromagnetic Interference in Hospitals*, 1994, pp. 7/1-7/4.
[13] G. D. O'Clock and J. B. Jarding, "Electrotherapeutic device/protocol design considerations for visual disease applications," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 2133-2136, doi: 10.1109/IEMBS.2009.5333966.