International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Prediction Using Machine Learning

Korla Surya Kiran

Student, Department of Information Technology, GMR Institute of Technology, Rajam, A.P

ABSTRACT:

Nowadays most of the people across the globe were losing their lives due to heart problems. to beat the reduction of loss of deaths, some accurate techniques are used to predict heart disease using machine learning algorithms. Supervised machine learning algorithms are powerful classification techniques to build prediction models that help diagnose the disease early. In this , HRFLM approach is employed ,it is combining the characteristics of Random Forest (RF) and Linear Model (LM). HRFLM proved to be quite accurate within the prediction of heart disease. Random Forest is employed to predict the disease with better accuracy compared to other algorithms. The random forest may be a classification algorithm consisting of many decisions trees. It uses bagging and have randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. There are many approaches in Classification and Regression. In this, random forest approach is employed which gives better accuracy which is more efficient than any other algorithms. Through this manner , training and predicting the model for heart condition problems in an efficient and effective manner.

INTRODUCTION

The study that is suggested in this paper focuses mostly on the various data used in heart condition prediction. The main component of the human body is the heart. In essence, it controls the flow of blood throughout our body. Any cardiac abnormality might exacerbate pain in other body areas. Any subtle alteration to the heart's normal operation might be categorised as a cardiac ailment. One of the main causes of most fatalities in the modern world is heart disease. Heart disease can be brought on by living a sedentary lifestyle, smoking, drinking alcohol, and eating a lot of fats, which can raise blood pressure. The World Health Organization estimates that more than 10 million people worldwide pass away each year as a result of heart disease. The only means of stopping heart-related ailments are a healthy lifestyle and early identification. Machine learning is frequently utilised in the medical industry for illness diagnosis, detection, and prediction. The main objective of this study is to give clinicians a tool for early heart disease detection. This will subsequently aid in providing patients with effective care and averting negative outcomes. In order to uncover hidden discrete patterns and analyse the provided data, ML plays a crucial role. Following data analysis, machine learning approaches aid in the early detection and prediction of cardiac disease. In order to predict cardiac disease at an early stage, this work offers a performance analysis of ML approaches called HRFLM (Hybrid Random Forest (RF) and Linear Model (LM)).

LITERATURE SURVEY

[1]. S. Mohan, C. Thirumalai and G. Srivastava, "Effective heart condition Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542-81554, 2019. In this paper, author propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of heart disease. He used Naive Bayes , Logistic Regression, Decision Tree, Random Forest and SVM. The proposed HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM).

[2]. Elwahsh, H., El-Shafeiy, E., Alanazi, S., & Tawfeek, M. A, "A new smart healthcare framework for real-time heart disease detection based on deep and machine learning", PeerJ Computer Science, 7, 2021.This paper is to develop a model using deep and machine learning techniques to predict the presence of heart disease. SVM, Logistic Regression, Decesion Tree, Naïve Bayes and Random Forest are used to analyse the data.Author compares the accuracy of these machine learning techniques and he finally got best accuracy for Random Forest.

[3]. R. Tao et al., "Magnetocardiography-Based Ischemic heart condition Detection and Localization Using Machine Learning Methods," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 6, pp. 1658-1667, June 2019.This study focused on developing a quick and accurate automatic ischemic heart disease detection/localization methodology.These features were categorized into three groups: time domain features, frequency domain features, and knowledge theory features. KNN, SVM, XGBoost are used and analyse the info ,best accuracy for XGBoost.

[4]. C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han and J. Yu, "Recursion Enhanced Random Forest with an Improved Linear Model (RERF-ILM) for heart condition Detection on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 59247-59256, 2020.This paper aims to seek out the key features of the prediction of cardiovascular diseases through the use of machine learning techniques.Decision Tree, SVM, Random Forest, Naïve Bayes, Neural Networks, K-Nearest Neighbour are wont to analyze the data. He got the simplest accuracy for

Random Forest.

[5]. F. A R. Buettner and M. Schunter, "Efficient machine learning based detection of heart condition ," IEEE International Conference on E-health Networking, Application & Services (HealthCom), 2019, pp. 1-6..This paper describes a way to detect possible heart disease using the Random Forests algorithm.This machine learning work contributes to healthcare and may detect heart disease on the basis of clinical data and test data from different patients.Author got better accuracy for Random Forest.

METHODOLOGY

In this paper, the problem mainly depends on classification. So we will use machine learning predictive models to find whether the person having Heart Disease or not , after predictive models are formulated using the training data.There are three types of classifications:

1. Binary Classification: Classification task with two possible outcomes.
2. Multi-class Classification: Classification with more than two classes.
3. Multi-label Classification: Classification task where each sample is mapped to set of target labels (more than one class).

Logistic Regression:

Logistic regression is one among the popular Machine Learning algorithms, which comes under the Supervised Learning technique. it's used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical variable . Therefore the result must be a categorical or discrete value. It are often either Yes or No, 0 or 1, true or False, etc. but rather than giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

$$Logistic\ function = \frac{1}{1+e^{-x}}$$

Logistic regression is employed for solving the classification problems

Decision Tree:

Decision Tree is a supervised learning technique that can be used for both classification and regression problems, but mostly it is preferred for solving classification problems .It is called Decision tree, because similar to a tree, it starts with a root node, which expands on further branches and constructs a tree-like structure .

By calculating how much accuracy each split will cost us, using a function . The split that costs least is chosen,this algorithm is recursive in nature as the groups formed can be sub-divided using same strategy.

Random Forest:

Random forest Supervised Machine Learning Algorithm that's used in Classification and Regression problems. It contains decision trees on different samples.Random forest may be a collection of decision trees. it's a common type of ensemble methods which is aggregate results from multiple predictors. Random forest additionally utilizes bagging technique that permits each tree trained on a random sampling of original dataset and takes the majority vote from trees. Compared to decision tree, it's better generalization but less interpretable, due to more layers added to the model.

K-Nearest Neighbor (KNN):

K-Nearest Neighbour is one among the simplest Machine Learning algorithms based on Supervised Learning technique.

K-NN may be a non-parametric and lazy learning algorithm. Non-parametric means there's no assumption for underlying data distribution i.e. the model structure is decided from the dataset. it's called Lazy algorithm because it does not need any training data points for model generation. All training data is employed in the testing phase which makes training faster and testing phase slower and costlier.

KNN works by finding the distances between a question and all the examples in the data, selecting the required number examples (K) closest to the query, then votes for the foremost frequent label (in the case of classification) or averages the labels (in the case of regression).

Support Vector Machines:

Support Vector Machine or SVM is one among the most popular Supervised Learning algorithms, which is employed for Classification as well as Regression problems. However, primarily, it's used for Classification problems in Machine Learning.

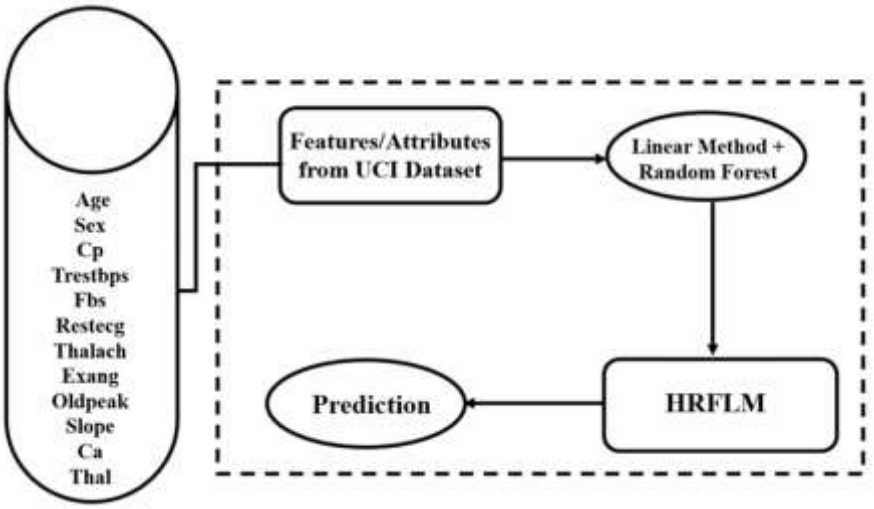
The goal of the SVM algorithm is to make the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is named a hyperplane.

RESULT AND DISCUSSION

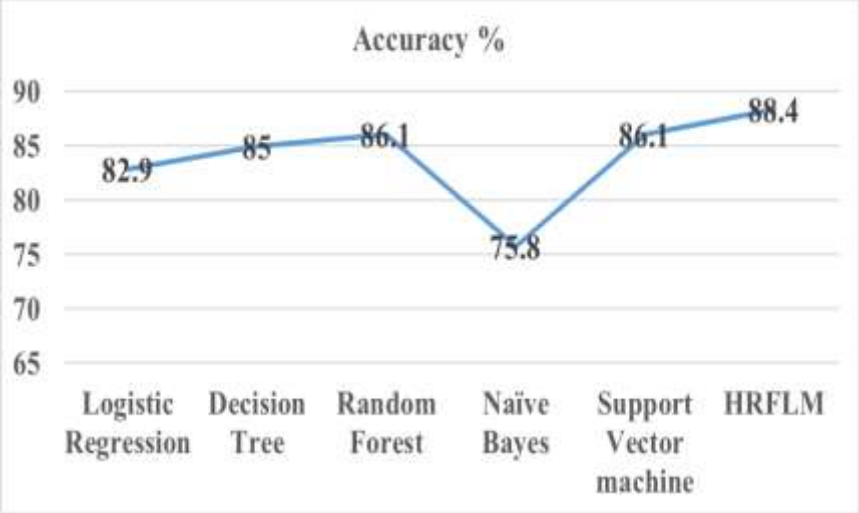
After analysing the data, train the model using machine learning techniques and get the accuracy for each model as follows in the table below.

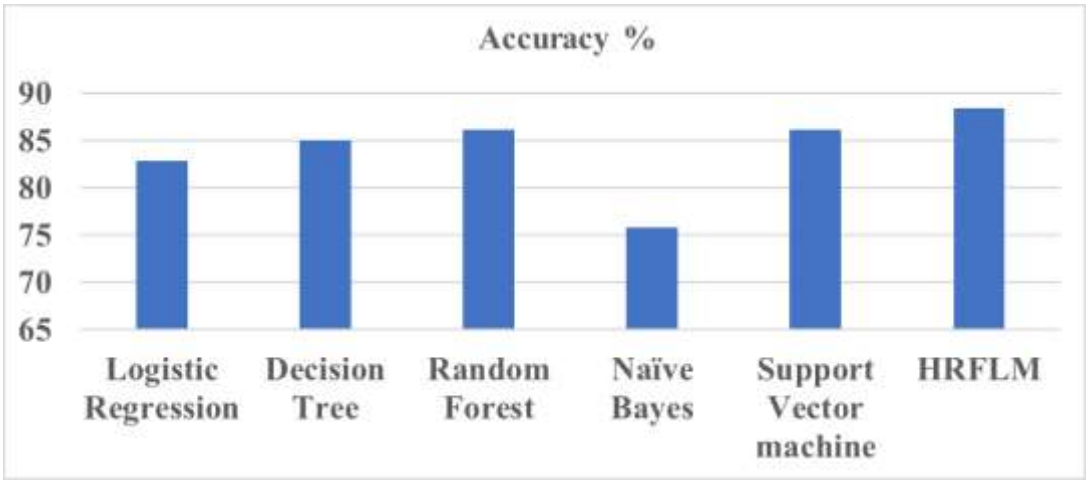
S. No	Modelcur	Accuracy %
1	Logistic Regression	82.9
2	Decision Tree	85
3	Random Forest	86.1
4	Naïve Bayes	75.8
5	SVM	86.1
6	HRFLM	88.4

The hybrid HRFLM approach is employed combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate within the prediction of heart disease.



HRFLM gives a best accuracy in the prediction of heart disease.





Conclusion

Machine learning techniques were used in this work to process raw data and provide a new prediction model towards heart disease.From the analysis conducted in this paper, different machine learning algorithm are applied on the data. Comparative study performed among the applied various techniques Logistic Regression ,Decision Tree, Random Forest, Naïve Bayes ,Support Vector machine and HRFLM. After analysing their approaches for Heart disease prediction. HRFLM(Hybrid Random Forest Linear Model) gives good result when compared to others. So HRFLM can be effectively gives best Accuracy 88.4%. The future course of this research can be performed with mixtures of machine learning techniques to better prediction techniques

References:

[1]. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542-81554, 2019.

[2]. Elwahsh, H., El-Shafeiy, E., Alanazi, S., & Tawfeek, M. A, "A new smart healthcare framework for real-time heart disease detection based on deep and machine learning", PeerJ Computer Science, 7, 2021.

[3]. R. Tao et al., "Magnetocardiography-Based Ischemic Heart Disease Detection and Localization Using Machine Learning Methods," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 6, pp. 1658-1667, June 2019.

[4]. C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han and J. Yu, "Recursion Enhanced Random Forest with an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 59247-59256, 2020.

[5]. R. Buettner and M. Schunter, "Efficient machine learning based detection of heart disease," IEEE International Conference on E-health Networking, Application & Services

[6]. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ''Analysis of neural networks based heart disease prediction system,'' in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239

[7]. P. K. Anooj, ''Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,'' J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan. 2012.

[8]. N. Al-milli, ''Backpropogation neural network for prediction of heart disease,'' J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[9]. M. S. Amin, Y. K. Chiam, K. D. Varathan, ''Identification of significant features and data mining techniques in predicting heart disease,'' Telematics Inform., vol. 36, pp. 82–93, Mar. 2019.

[10]. M. Gandhi and S. N. Singh, ''Predictions in heart disease using techniques of data mining,'' in Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manage. (ABLAZE), Feb. 2015, pp. 520–525