IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A SURVEY OF MACHINE LEARNING TECHNIQUES FOR PREDICTION OF HEART DISEASE

Dr. T.S. BASKARAN
Associate Professor& Research Supervisor,PG & Research Department of Computer Science, A Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,

Ms. B. MADHUVANTHI
Research Scholar,
PG & Research Department of Computer Science,
A. Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,

"Affiliated to Bharathidasan University, Tiruchirappalli-620024", TamilNadu, India.

Abstract:

The heart is one of the most important organ in Body. It helps in purifying the blood and sending it to all parts Body. Every year almost 26 million patients are affected with this heart disease. From the heart consultant and surgeon's point of view, it is complex to predict heart failure at the right time. Heart Disease is one of the major causes of death all around the world therefore early detection of Heart Disease can help reduce these rates. This research aims to use historical medical data to predict Heart Disease using Machine Learning technology. The discussed machine learning based on classification algorithms includes Support vector machine, Logistic regression, K-nearest neighbor, Naïve bays, Neural Network and Decision tree. The algorithms are compared based on features. We are working on the algorithm with the best accuracy. This will help the doctors to assist with the heart problem easily.

Keywords: Heart disease, Machine Learning, Prediction, Classification Technique,Accuracy.

I. Introduction

Heart disease is increasing daily in this modern world. Over the past decade, heart disease has been the leading factor of fatality globally. According to the World Health Organization, about 23.6 million people die primarily from cardiovascular disorders, with coronary heart disease and stroke accounting for 82 percent of these deaths. The ML methods are more precise and efficient than other methods without any human assistance [1].

Heart disease can occur due to the weakening of the heart muscle. Also, heart failure can be described as the failure of the heart to pump blood. Heart disease can occur due to insufficient blood supply to arteries. Heart disease can be detected using symptoms like high blood pressure, chest pain, hypertension, cardiac arrest, physical body weakness, feet are swollen, etc. Sometimes in heart disease, the infection also affects the inner membrane which is identified by symptoms like fever, fatigue, dry cough, and skin rashes. Causes of heart infection are bacteria, viruses, and parasites. Types of heart disease: heart block, hypertension, coronary artery disease, heart failure, heart infection, congenital heart disease, slow heart rate, stroke type heart disease, angina pectoris. Nowadays there are many automated techniques like data mining, machine learning, deep learning, etc. to diagnose heart disease. So, in this paper, we will brief introduction to machine learning techniques. In this, train the datasets using the machine learning repositories. There are some risk factors based on the heart disease is predicted. Risk factors are Age, Sex, Blood pressure, Cholesterol level, Family history of coronary illness, Diabetes, Smoking, Alcohol, Being overweight, Heart rate, and Chest Pain.

II. LITERATURE REVIEW

There are several works have been done associated with disease prediction systems using different machine learning algorithms in health Centres. Machine learning classification algorithms are vastly utilized in many fields to solve numerous problems. A field such as healthcare is considered a ridiculous machine learning field, where machine learning can be employed to tackle various medical decisions.

V.V. Ramalingam et Al,[2] proposed Heart disease prediction using machine learning techniques in which Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the healthcare industry and professionals in the diagnosis of heart-related diseases.

Avinash Golande et al,[3] proposed Heart Disease Prediction Using Effective Machine Learning Techniques in which Specialists utilize a few data mining strategies that are available to support the authorities or doctors in distinguishing heart disease. Usually utilized methodologies utilized are decision tree, k- closest, and Naïve Bayes. Other individual characteristic-based techniques used include packing calculation, partial thickness, continuous indifference regularization and neural systems, straight kernel self-organizing algorithm, and SVM (Bolster Vector Machine). Below are the details of the systems used in the exam.

Lakshmana Rao et al,[4] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc., Therefore, differentiating heart disease is difficult. A variety of systems in data mining and neuroscience have been used to detect the severity of heart disease in populations. The thought of CHD is frightening, and in addition, in this manner, the disease must be treated with caution. If not

recognized early, it can affect the heart or pass suddenly. Therapeutic science perspective Also, data mining can be used to discover different types of metabolomic machine learning processes, which causes past information experiments and models to be obtained without being personalized. Machine learning makes reasoning chronically informational.

Waigi at el., [5] A machine learning model is applied to the dataset taken from Kaggle. This dataset contains several parameters contributing to cardiac disease. It contains 70000 records and contains parameters like age, cholesterol, glucose, smoking, alcoholic habits, etc. The decision Tree model is used for training and predicting the risk of heart disease. Implementing the decision tree model on the cardio dataset, it is observed that the accuracy is not very good. Maybe Naïve Bayes classifier will be a better option for this dataset for predicting the risk of heart disease. Also to improve accuracy we will design a wristband that will continuously monitor pulse rate, body temperature, and blood flow rate. Adding these parameters to the dataset will improve accuracy.

Khan and Mondal [6] the heart disease using several machine learning algorithms including support vector machine (SVM), decision tree, k- nearest neighbors (kNN), logistic regression, naïve Bayes, random forest, and majority voting. The training and testing portion of each dataset is separated using holdout and cross-validation methods. The performance of different algorithms for three datasets is evaluated in terms of testing accuracy, precision, recall, and F1- score. It is shown here that majority voting is a combination of logistic regression, SVM, and naïve Bayes.

Shorewall [7] Learning techniques such as K-Nearest Neighbors, Binary Logistic Classification, and Naive Bayes will be used with a cross-comparative study. K-Fold's validation will be employed as well to create randomness in the data and to look at the consistency of the results produced by the models. Furthermore, hybrid models are also explored using ensemble techniques such as bagging, boosting, and stacking. These ensemble techniques are cross-compared to the results of the original base classifiers. These algorithms are tested on the 'Cardiovascular Disease Dataset,' The stacked model involving KNN, randomforest classifier, and SVM proved to be the most effective with an accuracy of 75.1%.

Our and ElSeddawy [8] apply the four-cross-validation techniques (holdout, k-fold cross- validation, stratified k-fold cross-validation, and repeated random) with the eight data mining, classification techniques (Linear Discriminant Analysis, Logistic regression, Support Vector Model, KNN, Decision Tree, Naïve Bayes, Random Forest, and Neural Network) to improve the accuracy of heart disease prediction and select the best prediction models. It analyzes these techniques on a small and large dataset collected from different data sources like Kaggle and the UCI machine-learning repository. The evaluation metrics like accuracy, precision, recall, and F-measure were used to measure the performance of prediction models. Experimentation. The best models will be recommended to the physicians in business organizations to help them predict heart disease in employees into one of two categories, cardiac and non-cardiac, at an early stage. The early detection of heart diseases in employees will improve productivity in thebusiness organization.

Ramesh T .,[9] Machine Learning methods significantly outperformed statistical techniques in this research. This article proves the studies of various researchers which suggest that the use of ML models is the best choice to predict and classify heart disease even if the database is more diminutive. Various performance parameters, i.e., precision, F1 score, accuracy, and recall, have been compared for the entire Machine Learning classification techniques on the UCI online heart disease data set. The KNN classification algorithm outperformed the fourteenavailable parameters.

Manal Makram [10] machine learning models to classify cardiovascular diseases. The heart disease dataset was obtained from Egypt's Ain Shams University, Specialized Hospital for the inpatient department's Coronary Care Unit. A comparative analysis of several classifiers indicates that the neural network achieves the best results with accuracy.

Dhruva R. Rinku .,[11] Machine Learning (ML) models to predict the HD based on the existing symptoms of the patients. The dataset from the UPI repository is used to evaluate the performance of the proposed models. The various parameters namely precision (p), recall (r), and accuracy (a) are used to evaluate the performance measures of the ML models. Observing the results concluded that, the Random Forest model outperformed the other models such as XGBoost, Decision Tree, and traditional Neural Network model regarding the prediction accuracy concerning the UCI dataset.

This paper presents a survey of various models based on such algorithms and techniques and their survey analysis report in Table 1. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), and ensemble models are found very popular among the researchers and systems have been applied to different clinical datasets to robotize the investigation of huge and complex information. Several scientists, as of late, have been applying a few Machine Learning algorithms, and techniques have been applied to several medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the healthcare industry and professionals in the diagnosis of heart-related diseases. Strategies to enable the well-being of to mind industry and the experts in the analysis of heart-related sicknesses. This paper presents a review of different models dependent on such calculations and methods and analyzes their exhibition. Models in light of directed learning calculations, for example, Support Vector Machines (SVM), K- Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), Neural Network and group models are discovered extremely well-known among the scientists.

Table 1Survey analysis report

YEAR	AUTHOR	PURPOSE	TECHNIQUE USED	ACCURACY
2018	V.V.Ramalingam [2]	Heart disease prediction using machinelearning techniques	1) Naïve Bayes 2) Support Vector Machine 3) K – Nearest Neighbour 4) Decision Tree 5) Random Forest 6) Ensemble	SVM has more accuracy than other techniques.
2019	Avinash Golande [3]	Heart Disease Prediction Using Effective Machine Learning Techniques	1)Decision Tree2)KNN 3)K-mean clustering 4)Adaboost	Decision tree (86.60% Accuracy)
2019	A. Lakshmanarao[4]	Machine Learning Techniques For Heart Disease Prediction	1)Random Over sampling2)Synthetic Minority Oversampling 3)Adaptive synthetic sampling approach	For random oversampling, SVM given the best accuracy of **82.30%.** For Synthetic Minority Oversampling, Random Forest
				given the best accuracy of **91.3%** For Adaptive Synthetic sampling, RandomForest (**90.3%** Accuracy)
2020	Waigi at el., [5]	Predicting the risk of heart disease using advanced machine learning approach.	1. Decision Tree 2.Naïve Bayes Classifier	72.77% (decision tree)
2020	Khan and Mondal [6]	Data-Driven Diagnosis of Heart Disease.	1.logistic regression 2.SVM 3. Naïve Bayes	72.72% (logistic regression)
2021	Shorewall [7]	Early detection of coronary heart disease using ensemble techniques.	Ensemble techniques such as bagging, boosting, and stacking.	75.1% (The stacked model involving KNN, random forest classifier, andSVM)

2021	Shimaa Ouf and Ahmed I. B. ElSeddawy [8]	A proposed paradigm for intelligent heart disease prediction system using data mining techniques	1.Linear Discriminant Analysis, 2.Logistic regression, 3.Support Vector Model, 4. KNN, 5. Decision Tree, 6.Naïve Bayes, 7.Random Forest, 8.Neural Network	89.01%(random forest classifier)
2022	Ramesh T .,[9]	Predictive analysis of heart diseases with machine learning approaches	1. Naive Bayes 2. SVM 3.Logistic regression 4. Decision Tree Classifier 5. Random Forest, 6.K-Nearest Neighbor	94% KNN classification algorithm
2022	Manal Makram [10]	Machine learning approach for Diagnosis of heart diseases	1.SVM, 2.Logistic Regression, 3.KNN, 4. Random Forest, 5. K-Means, 6.Naïve Bayes, 7.Decision Tree, 8. Neural Network	Neural Network (96%)
2023	Dhruva R. Rinku .,[11]	Performance Comparison Analysis of Predicting the Heart Diseases using Machine Learning Algorithms	1. Random Forest 2.XGBoost 3.Decision Tree 4.Neural Network	Random Forest (96%)

III. CONCLUSION AND FUTURE WORK

To summarized different types of machine learning algorithms for prediction of heart disease. To elaborate various machine learning algorithms and worked towards finding the best algorithm by analysing their features. Every algorithm has given different result in different situations. Further it is analysed only marginal accuracy is achieved for predictive model of heart disease and hence more complex models are needed to increase the accuracy of predicting the early heart disease. In future we will propose methodology for early prediction of heart disease with high accuracy and minimum cost and complexity.

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, doi: 10.1109/ACCESS.2019.2923707.

[2] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, "Heart disease prediction using machine learning techniques: A survey," Int. J. Eng. Technol., vol. 7, no. 2.8 Special Issue 8, pp. 684–687, 2018, doi: 10.14419/ijet.v7i2.8.10557.

[3] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[4] A. Lakshmanarao,Y.Swathi, P.Sri Sai Sundareswar, "Machine Learning Techniques for Heart Disease Prediction," 2023 Somaiya Int. Conf. Technol. Inf. Manag. SICTIM 2023, no. October, pp. 28–33, 2023, doi: 10.1109/SICTIM56495.2023.10104919.

[5] R. Waigi, S. Choudhary, P. Fulzele, and G. Mishra, "Predicting the risk of heart disease using advanced machine learning approach," Eur. J. Mol. Clin. Med., vol. 7, no. 7, pp. 1638–1645, 2020, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L2010514072&from=export

[6] M. I. H. Khan and M. R. H. Mondal, "Data-Driven Diagnosis of Heart Disease," Int. J. Comput. Appl., vol. 176, no. 41, pp. 46–54, 2020, doi: 10.5120/ijca2020920549.

[7] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," Informatics Med. Unlocked, vol. 26, p. 100655, 2021, doi: 10.1016/j.imu.2021.100655.

[8] F. Programmatic and E. Impact, "A proposed paradigm for intelligent heart disease prediction system using data mining techniques" vol. 76, no. 30, pp. 1–40, 2011.

[9] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "Predictive Analysis of Heart Diseases With Machine Learning Approaches," Malaysian J. Comput. Sci., vol. 2022, no. Special Issue 1, pp. 132–148, 2022, doi: 10.22452/mjcs.sp2022no1.10.

[10] M. Makram, N. Ali, and A. Mohammed, "Machine Learning Approach for Diagnosis of Heart Diseases," MIUCC 2022 - 2nd Int. Mobile, Intelligent, Ubiquitous Comput. Conf., pp. 69–74, 2022, doi: 10.1109/MIUCC55081.2022.9781735.

[11] D. R. Rinku, S. D. Gogula, S. Prathipati, P. V. Ramana Murthy, R. K. Yarava, and U. D. Kosuri, "Performance Comparison Analysis of Predicting the Heart Diseases using Machine Learning Algorithms," 2023 4th Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2023 - Proc., no. August, pp. 1272–1277, 2023, doi: 10.1109/ICESC57686.2023.10193657.