International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228656

Heart Disease Prediction Using Machine Learning Algorithm

Mahendra Patidar[1], Prof. Vinod Mahor[2]
[1]Millennium Institute of Technology and Science, Bhopal, Madhya Pradesh, India
[2]Assistant Professor, Millennium Institute Technology and Science, Bhopal, Madhya Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 366-373

Article History
Accepted: 15 Nov 2022
Published: 03 Dec 2022

ABSTRACT

Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparison to the previously used classifier such as naive bayes etc. So a quiet significant amount of pressure has been lift off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format

Keywords : Heart Disease, KNN

I. INTRODUCTION

"Machine Learning is a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data" [1]. Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset. We can use that knowledge in our project of HDPS as it will help a lot of people. The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc. A dataset is selected from the UCI repository with patient's medical history and

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

attributes. By using this dataset, we predict whether the patient can have a heart disease or not. To predict this, we use 14 medical attributes of a patient and classify him if the patient is likely to have a heart disease. These medical attributes are trained under three algorithms: Logistic regression, KNN and Random Forest Classifier. Most efficient of these algorithms is KNN which gives us the accuracy of 88.52%. And, finally we classify patients that are at risk of getting a heart disease or not and also this method is totally cost efficient.

II. LITERATURE REVIEW

As opposed to classical ML, which depends exclusively on the loss of accuracy of the model, the author of paper [1] offered a model with a superior approach that gains more accuracy and confidence. We tested our methodology using the Pima Indian Diabetes dataset (PIDD) and breast cancer, and with ensemble learning, we grow confident in the dependability of our models. For a very long time, the diagnosis of any disease has relied heavily on the prediction and early detection of diseases. Algorithms for machine learning (ML) have shown to be quite effective at identifying diseases and making healthcare decisions. Even though the majority of ML algorithms had outstanding accuracy, domain adaptability and resilience remain major issues. [1] While certain algorithms do well on some datasets, they struggle on others. Updated ML algorithms are required since the performance of these algorithms can frequently alter in the future due to data variance. In paper [2], the author suggests a methodology for predicting HD and T2D in chronic diseases. The outlier identification method for the proposed study was random forest mixed with DBSCAN, while the data balancing method was SMOTE-ENN. The model was developed using three HD datasets (Statlog and Cleveland) and one T2D dataset (NHIS Korea), and the output was compared with several machine learning (ML) algorithms, such as GNB, LR, MLP, DT, and SVM. In this study, k-fold (10) cross-validation and a number of performance metrics, such as accuracy, precision, f-measure, and recall, are used to assess the model's performance. The accuracy rates for the Statlog HD dataset, Cleveland HD dataset, and NHIS T2D dataset for the model we suggested are, respectively, 97.63%, 97.69%, and 94.85%, outperforming other classification algorithms and earlier studies. When the state of HD and T2D are recognised, using the proposed model could raise the likelihood that the worst case scenario would be avoided and assist people in acting swiftly and precisely. The leading chronic disease and leading cause of disability and death globally is heart disease (HD). Along with HD, type 2 diabetes (T2D) is one of the most fatal illnesses that can have catastrophic effects if undiagnosed and mistreated. The best methods to prevent HD and T2D are HD and T2D predictions. In order to assist people in avoiding the worst situations, early HD and T2D predictions are crucial.

In a paper [3], the author conducted a comparative examination of the models based on different performance criteria (accuracy, precision, and recall). Last but not least, ensemble learning is utilised to predict death from HF with a 90% success rate, as demonstrated by test data. High blood pressure in the body's arteries characterises the chronic disease known as hypertension. The effect is that the heart must work harder to maintain the body's regular blood flow. It is a major contributor to future fatal and non-cardiovascular illnesses, stroke, and renal failure risk factors. The long-term risk of older participants (over 50 years old) receiving a hypertension diagnosis is predicted using machine learning (ML) in this article[3]. Our goal is to develop models that are highly sensitive to recognising individuals who are at risk so that, after the appropriate actions, hypertension won't develop or manifest itself in the future. The proposed technique compares two ML models (Decision Tree and Naive

Bayes) based on Precision, Recall, F-Measure, Accuracy, and Area Under Curve. Two different class balancing strategies are also taken into consideration (AUC).

In Paper [4], multiple machine learning (ML)-based methods are used to a dataset of heart disease to forecast the mortality of HF patients. In India as well as the rest of the globe, 17.9 million fatalities every year account for 31% of all deaths. One of the most fatal diseases that is frequently brought on by CVDs is HF. It is essential in medicine to make the right decisions quickly when treating patients. Given the enormous amount of data that the healthcare industry generates [4], ML systems are crucial for predicting CVD.

The dataset, used in paper [5] which consists of 303 rows and 14 attributes, is from Kaggle. Logistic Regression, NB, K-NN, SVM, Multi-Layer Perceptrons, Artificial Neural Networks, Decision Tree, Random Forest, XG Boost, and Cat Boost are the algorithms utilised in the model. According to WHO statistics, cardiac disorders are responsible for close to 1 crore 20 million deaths annually. Heart disease and cardiovascular illness have historically had a significant impact on the medical sector, making them extremely dangerous and widespread. Although it would take a lot of expertise and time, it is not possible to forecast cardiac problems or CD or to monitor a patient around-the-clock. [5]Heart disease diagnosis and treatment are highly difficult, especially in developing or underdeveloped nations. Inadequate medical care or a delay in treatment of a disease can also result in a person's death. Researchers frequently use the wealth of data from the medical sector to create new science and technologies aimed at reducing the number of heart-related deaths. There are numerous data mining and machine learning approaches and algorithms that can be used to retrieve data from databases and use that data to forecast cardiac problems with high accuracy.

The author of paper [6] suggested a model that can assist physicians in determining a patient's current heart disease condition. As a result, by starting early therapy, heart disease-related mortality can be avoided. The most common cause of death in the world is cardiovascular disease (CVD). Based on clinical data, a machine learning (ML) system can predict CVD in the early stages to reduce death rates. Numerous studies have recently used various machine learning techniques to detect CVD or determine the severity of the patient's condition. Despite the positive outcomes of these studies, none of them concentrated on using optimization techniques to enhance the ML model's performance for CVD detection and severity-level classification. The Synthetic Minority Oversampling Technique (SMOTE), six different ML classifiers to determine the patient's status, and Hyperparameter Optimization (HPO) to determine the ideal hyperparameter for ML classifier in conjunction with SMOTE are all used in this study to provide an efficient method for handling the imbalance distribution issue. [7] The model with all features was constructed and tested on two open datasets. The findings demonstrate that SMOTE and Extra Trees (ET) optimised utilising hyperband exceeded state-of-the-art studies by obtaining 99.2% and 98.52% in CVD detection, respectively, outperforming other models. Additionally, utilising the Cleveland dataset, the created model converged to a severity classification accuracy of 95.73%.

As part of investigation, in paper [7] also looked for relationships between the different database variables in order to use those relationships to forecast the likelihood of developing heart disease. As a decision support system, this model might be helpful to the medical personnel at their clinic. Cardiovascular problems are becoming more prevalent, according to a recent WHO research. As a result, we can deduce that around 17.9 million people pass away each year. Early diagnosis and treatment are becoming more and more challenging as the population grows. But thanks

to recent technological developments, machine learning methods have sped up the health sector's research. Therefore, the aim of this study is to build an ML model for heart disease prediction based on related parameters.

Different machine learning methods, including decision trees, K-Nearest Neighbor, and AdaBoost, are utilised in Paper [9]. Predicting the sickness is the main objective of this research. The accuracy of each algorithm is assessed when it has been fully implemented. The Kaggle website offers a free download of the dataset. The heart is an essential part of all living things. The most lethal disorder in the world is heart disease, which occurs when the heart is unable to deliver enough blood to other parts of the body. Heart-related disease prediction and diagnosis demand better precision and accuracy because even a slight error could result in exhaustion or even death. Heart-related deaths occur often, and the rate of occurrence is rising. The old approach of detecting heart disease is not widely regarded as being accurate. It is essential to establish a predictive system to increase awareness of illnesses in order to address the problem. Artificial intelligence (AIdiscipline)'s of machine learning provides a famous service to forecast all kinds of events using training data from observed natural phenomena.

In this [10], the author suggests and executes a cost-effective and cheap ML model that can, given a patient's clinical record, determine early on whether they have heart disease or not. The proposed model, which uses the Cleveland dataset, combines many machine learning (ML) classifiers, including K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), and Naive Bayes (NB). The model uses data preparation techniques such feature extraction, standardisation, and classification. In this study, it is found that, when compared to all other classifiers including the Decision Tree, the Nave Bayes classifier provides the best prediction with an accuracy of 86.88%.

The author of study [11] took heart illness into account for experiments, and experimental evaluation determined the forecast for classification categories. The mean test score (m) is 20 when the number of decision trees (M) with samples (MS), leaf nodes (ML), and learning rate (I) are calculated as MS=20, ML=3, and I=0.1. Health conditions have become severely damaging to human life as a result of various dehydrated foods and disruptions to the workplace environment. For the major purposes of deterrent, recognition, and treatment, accurate disease prediction and diagnosis becomes a more important and difficult issue. We therefore presented the Medical Things (MT) and machine learning models based on the aforementioned difficulties to address the healthcare issues with suitable services in disease supervision, forecasting, and diagnostics. We created a prediction framework using machine learning techniques to obtain various classification groups for diseases that were anticipated. To reduce the complexity of the data, the framework was created using a decision tree and a fuzzy model.

Heart disease is currently the leading contributor to fatalities. Although early detection of cardiac illness can save lives, if the diagnosis is made too late, it could result in fatalities. Early heart disease diagnosis is an extremely challenging endeavour. The earlier forecast reveals the patient's heart state so that effective treatment can be given. Machine learning (ML) techniques can be used to resolve difficult and nonlinear situations. [12] Various machine learning (ML) techniques can be combined to improve categorization and prediction accuracy (ensemble models). These include K-Nearest Neighbor, Random Forest, and Decision Tree. ML approaches have been used for ensemble models. The ensemble model makes a prediction using the majority voting method.

The study of machine learning is essential for modern computer scientists. A crucial link exists between artificial intelligence and machine learning-based research. This study's authors employed machine learning to evaluate a person's likelihood of having cardiac illness. Cardiovascular diseases (CVDs) are common and are the leading cause of death worldwide. By taking into account certain indicators like chest discomfort, cholesterol levels, age, and other parameters, machine learning may be used to identify whether a person has a cardiovascular disease. Cardiovascular disease diagnostics can be made simple using classification algorithms based on supervised learning, a type of machine learning. Algorithms like K-Nearest Neighbor (KNN), Random Forest, Logistic Regression, Decision Tree Classifier, SVM, Naive Bayes, and Gradient Boosting Classifier are used to categorise people with heart disease. The supervised machine learning methods K-Nearest Neighbor (K-NN) and Random Forest are extensively used in this research.

III. Methodology

Various machine learning algorithms are analysed in this paper, including K nearest neighbours (KNN), logistic regression, and random forest classifiers, which can help practitioners or medical analysts diagnose heart disease accurately. This paperwork entails looking over journals, papers that have been published, and recent cardiovascular disease statistics. The proposed model has a framework provided by methodology [13]. The methodology is a procedure that involves stages that convert provided data into acknowledged data patterns for the consumers' awareness. The proposed methodology (Figure 1) consists of three steps: data collection in the first stage, substantial value extraction in the second stage, and data exploration in the third stage of pre-processing. Depending on the algorithms utilised, data pre-processing addresses missing values, data cleaning, and normalisation [13]. Pre-processed data are then classified using a classifier. The classifiers employed in the proposed model are KNN, Logistic Regression, and Random Forest Classifier. The suggested model is then put into practise, and its accuracy and performance are assessed using a variety of performance indicators. Using several classifiers, an efficient Heart Disease Prediction System (EHDPS) has been created in this model. For prediction, this model incorporates 13 medical variables, including age, sex, blood pressure, cholesterol, blood sugar during fasting, and chest pain [13s].

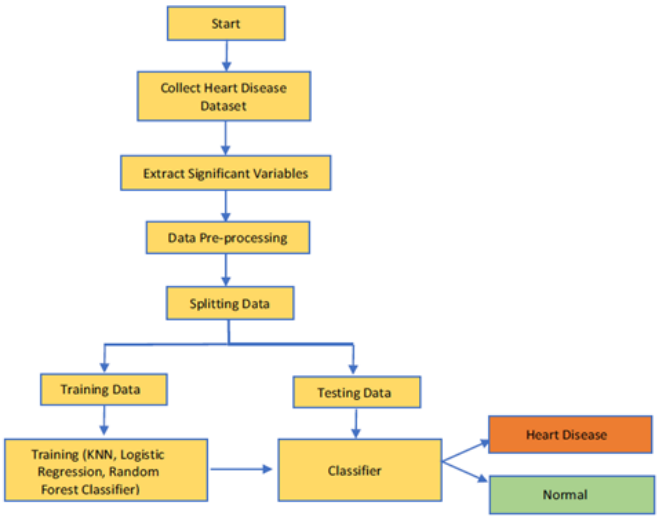


Figure 1: Proposed Model [13]

IV. Simulation and Result

In a simulation and result the data set from kengle website has been taken. Simulation has been done on Jupiter notebook. From these results we can see that although most of the researchers are using different algorithms such as SVC, Decision tree for the detection of patients diagnosed with Heart disease, KNN, Random Forest Classifier and Logistic regression yield a better result to out rule them [23]. The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from

previous researches. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took. Our project also tells us that Logistic Regression and KNN outperforms Random Forest Classifier in the prediction of the patient diagnosed with a heart Disease. This proves that KNN and Logistic Regression are better in diagnosis of a heart disease. The following 'figure 2', 'figure 3', 'shows a plot of the number of Different feature of data set and patients that are been segregated and predicted by the classifier depending upon the age group, Resting Blood Pressure, Sex, Chest Pain:

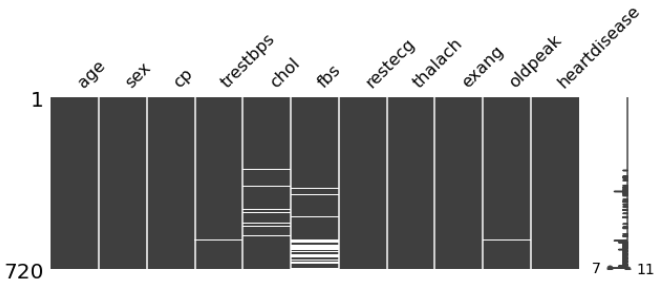
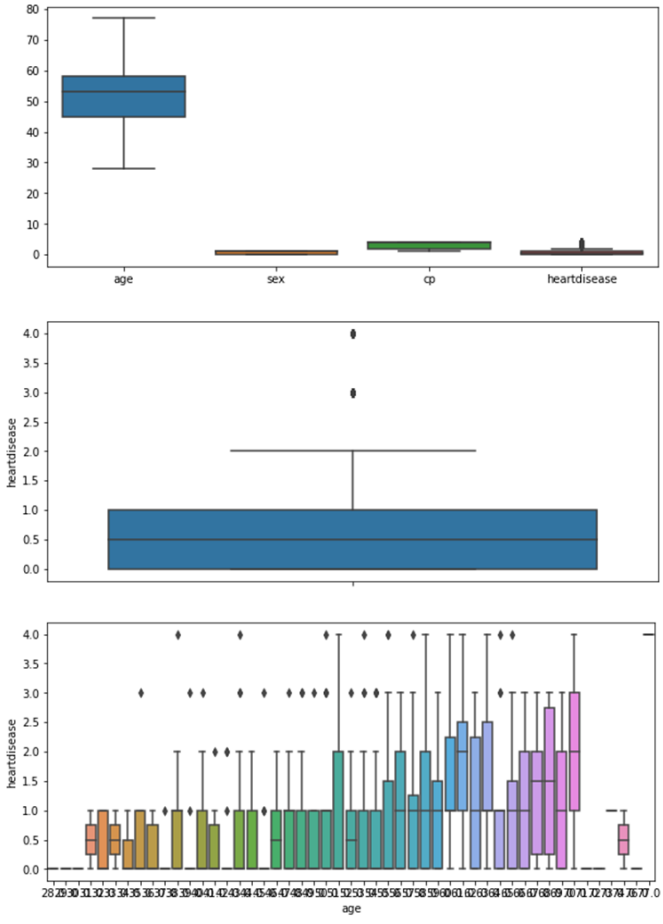


Figure 2 : Different Features of Data Set (Graph representation)



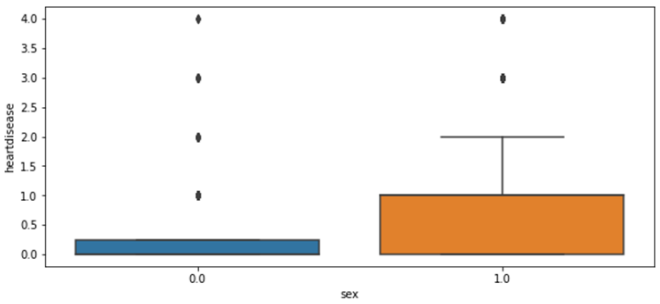


Figure 3 : Heart disease Prediction on Basis of Age and Sex

V. CONCLUSION AND FUTURE WORK

Three ML classification modelling techniques have been used to create a model for the detection of cardiovascular disease. By extracting the patient medical history that results in a fatal heart illness from a dataset that contains patients' medical history such as chest pain, sugar level, blood pressure, etc., this method predicts persons with cardiovascular disease. Based on clinical information about a patient's prior heart disease diagnosis, this heart disease detection system helps the patient. The proposed model was constructed using the methods of KNN, Random Forest Classifier, and Logistic Regression [22]. Our model has an accuracy rate of 87.5%. The likelihood that the model will correctly identify whether a specific person has heart disease or not increases with the use of more training data [9]. These computer-assisted tools allow us to anticipate the patient quickly, more accurately, and at a significantly lower cost. We can work with a variety of medical databases since machine learning techniques are more advanced and can anticipate outcomes better than humans, benefiting both patients and medical professionals. As a result, our research aids in the prediction of patients who are diagnosed with heart problems by cleaning the dataset and using logistic regression and KNN to achieve an accuracy of an average of 87.5% on our model, which is better than the prior models' accuracy of 85%. Additionally, it is established that KNN's accuracy of 88.52% is the greatest of the three

methods we utilised. According to "Figure 6," 44% of those reported in the dataset have cardiac disease.

VI. REFERENCES

[1]. Kanak Kumar; Kaustav Chaudhury; Suman Lata Tripathi, "Future of Machine Learning (ML) and Deep Learning (DL) in Healthcare Monitoring System," in Machine Learning Algorithms for Signal and Image Processing , IEEE, 2023, pp.293-313, doi: 10.1002/9781119861850.ch17.

[2]. N. L. Fitriyani, M. Syafrudin, G. Alfian, C. -k. Yang, J. Rhee and S. M. Ulyah, "Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest," 2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS), Manama, Bahrain, 2022, pp. 289-294. doi: 10.1109/ICETSIS55481.2022.9888806

[3]. E. Dritsas, S. Alexiou and K. Moustakas, "Efficient Data-driven Machine Learning Models for Hypertension Risk Prediction," 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Biarritz, France, 2022, pp. 1-6. doi: 10.1109/INISTA55318.2022.9894186

[4]. S. Kedia and M. Bhushan, "Prediction of Mortality from Heart Failure using Machine Learning," 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 2022, pp. 1-6. doi: 10.1109/ICEFEET51821.2022.9848348

[5]. K. Vayadande et al., "Heart Disease Prediction using Machine Learning and Deep Learning Algorithms," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 393-401. doi: 10.1109/CISES54857.2022.9844406

[6]. A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," in IEEE Access, vol. 10, pp. 79974-79985, 2022.

[7]. A. Srivastava and A. k. Singh, "Heart Disease Prediction using Machine Learning," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 2633-2635. doi: 10.1109/ICACITE53722.2022.9823584

[8]. Sanyal, D. Das, S. K. Biswas, M. Chakraborty and B. Purkayastha, "Heart Disease Prediction Using Classification Models," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6. doi: 10.1109/INCET54531.2022.9824651

[9]. K. Bhuyan, T. Arun Sai, M. Charan, K. Vignesh Chowdary and B. Brahma, "Analysis of classification based predicted disease using machine learning and medical things model," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-6. doi: 10.1109/ICAECT54875.2022.9807903

[10]. Cathey et al., "Using Machine Learning to Predict Development of Heart Failure, during Post-Acute COVID-19, by Race and Ethnicity," 2022 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2022, pp. 276-281. doi: 10.1109/SIEDS55548.2022.9799382

[11]. G. Thilagavathi, S. Priyanka, V. Roopa and J. S. Shri, "Heart Disease Prediction using Machine Learning Algorithms," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022,

pp. 494-501. doi: 10.1109/ICAAIC53929.2022.9793107

[12]. P. Kundu et al., "A Comparative Study on Prediction of Heart Disease and Classifiers Suitable Analysis," 2022 Interdisciplinary Research in Technology and Management (IRTM), Kolkata, India, 2022, pp. 1-5. doi: 10.1109/IRTM54583.2022.9791722

[13]. Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072DOI 10.1088/1757-899X/1022/1/012072

Cite this article as :

Mahendra Patidar, Prof. Vinod Mahor, "Heart Disease Prediction Using Machine Learning Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 366-373, November-December 2022. Available at doi : https://doi.org/10.32628/CSEIT228656
Journal URL : https://ijsrcseit.com/CSEIT228656