WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Heart Disease Prediction Using Machine Learning

Prof. Roshan Kolte[1,a)], Pranjal Waghmare[2,b)], Dip Chafale[3,c)], Jayesh Parkhi[4,d)], Khemraj Pusnake[5,e)], Ujwal Kamdi[6,f)]

[1,2,3,4,5,6]*K.D.K. College of Engineering, Information Technology, RTMNU, Nagpur, Maharashtra, India*

[a)] roshan.kolte@kdkce.edu.in, [b)] pranjalawaghmare.it21d@kdkce.edu.in, [c)] dipdchafale.it20f@kdkce.edu.in, [d)] jayeshvparkhi.it20f@kdkce.edu.in, [e)] khemrajdpusnake.it20f@kdkce.edu.in, [f)] ujwalvkamdi.it20f@kdkce.edu.in

ABSTRACT

Machine learning has the potential to be a critical tool in the diagnosis and prognosis of heart illnesses, loco-motor disorders, and other conditions. Due to its ability to identify patterns in data, machine learning applications in the medical field have grown. Diagnosticians can decrease misdiagnosis by using machine learning to categorize the occurrence of cardiovascular illness. If foreseen well in advance, such information might give physicians valuable insights that allow them to modify their diagnosis and treatment plan according to each patient. Our goal is to use machine learning algorithms to predict potential heart diseases in humans. We will compare various classifiers such as decision trees, Naïve Bayes, SVM, Random Forest, and Logistic Regression in this project. We will also propose an ensemble classifier that combines strong and weak classifiers to perform hybrid classification. Because this classifier can have multiple samples for training and validating data, we will analyze both the existing and proposed classifiers to provide better accuracy and predictive analysis. This method leverages previously completed patient records to forecast a new one at an early stage, sparing lives. In order to lessen the number of people who die from cardiovascular diseases, our research will create a model that can accurately forecast cardiovascular disorders. For medical professionals, predicting and detecting heart disease has always been a crucial and difficult undertaking. Heart disease can be treated with costly medicines and surgeries provided by hospitals and other facilities. Therefore, early detection of cardiac disease will be beneficial to individuals worldwide, enabling them to take the appropriate action before the condition worsens.

INTRODUCTION

Worldwide, machine learning is applied in a wide number of contexts. The health care sector is not an exception. Machine learning has the potential to be a critical tool in the diagnosis and prognosis of heart illnesses, locomotor disorders, and other conditions. We have analyzed data points using several machine learning models in order to produce plausible diagnoses and forecasts. These trained models are now available to the public and health professionals for use in enhanced risk assessment, recommendations, and complementary care. This technology makes it simple to weed out some patients and focus resources and attention on those who are more in need. It also helps us identify people who are in excellent health so that those who are truly suffering can receive the care they need without wasting time or money. Heart disease prognosis is one of the key ideas in the modern world that is influencing how society views health. The basic idea is to use the Random Forest algorithm to determine the age group and heart rate. Our project describes how a system's heart rate and condition are estimated using user-provided inputs, including blood pressure and many more. When compared to other algorithms, this is a much better method because RFA implementation improves user experience and yields accurate results. This is useful in a variety of contexts, including early disease prediction, where input is given to determine heart rate in relation to a given medical condition.

HISTORY OF MACHINE LEARNING IN HEART DISEASE PREDICTION

The utilization of machine learning technologies in healthcare has numerous potential applications, including enhanced patient data, medical research, diagnosis, and treatment, as well as cost reduction and improved patient safety. Here's a summary of some advantages healthcare professionals can expect from machine learning applications in the field:

Improving diagnosis: Medical practitioners can use machine learning to create more advanced diagnostic tools for analyzing medical images. For instance, a machine learning algorithm can be used to medical imaging to find patterns in images that suggest a specific disease through pattern recognition. With the use of this kind of machine learning algorithm, physicians may be able to diagnose patients more quickly and accurately, improving patient outcomes.

Developing new treatments: Healthcare institutions and pharmaceutical companies can also use a deep learning model to find pertinent information in data that may help with drug discovery, the creation of new medications by pharmaceutical companies, and the development of novel disease treatments. For instance, clinical trial data and medical research could be analyzed using machine learning in the healthcare industry to uncover previously

undiscovered medication side effects. In clinical trials, this kind of machine learning for healthcare could lead to improvements in drug discovery, patient care, and the efficacy and safety of medical procedures.

Reducing costs: Healthcare organizations can use machine learning technologies to increase healthcare efficiency, which may result in cost savings. In the healthcare industry, machine learning, for instance, might be utilized to create more effective algorithms for appointment scheduling and patient record management. The healthcare system may be able to save time and money by using this kind of machine learning to help with repetitive tasks.

Improving Care: Medical practitioners can use machine learning in healthcare to raise the standard of patient care. The healthcare sector may employ deep learning algorithms to create systems that proactively monitor patients and notify electronic health records or medical devices when a patient's condition changes. Making sure patients receive the right care at the right time may be made easier with the use of machine learning for data collection.

While the potential of machine learning to provide healthcare is still being explored, its applications are already having a positive effect on the field. Machine learning will play a bigger role in healthcare in the future as we try to make sense of the ever-expanding clinical data sets.

LITERATURE SURVEY

LIAQAT ALI et al. [4] proposed a model that combines two techniques: deep neural network (DNN) and the X2 statistical method. Features are refined using the X^2 statistical method. classification is carried out by a deep neural network (DNN). Their research has made use of the Cleveland collection. That dataset contains 303 instances, some of which are 6 of the 297 have missing data, and the other 297 have no missing data. information. 207 of the 297 instances are utilized as training data. And testing data is derived from the remaining 90. This particular model produces superior outcomes to traditional ANN models. which were there before. Because of this, utilizing this suggested model, achieving a 93.33% classification rate. precision with DNN. It surpasses by 3.33% that of traditional ANN model.

A data mining model was created by Dr. Kanak Saxena et al. [12] to accurately predict heart disease. It primarily assists medical professionals in making effective decisions within the specified parameters. The Cleveland dataset from UCI was used by the author, who also included attributes like sex, age, chest pain, serum cholesterol, fasting blood sugar, and resting blood pressure. Additionally, the datasets have been split into two sections: one for training and the other for testing. To find accuracy, they have employed a ten-fold approach.

An expert system based on two support vector machines (SVM) was proposed by AWAIS NIMAT et al. [1] to effectively predict heart disease. Both of these SVMs serve a purpose: the first is used for prediction, and the second is used for feature removal. Additionally, they optimized the two approaches using the hybrid gird search algorithm, or HGSA. They have obtained 3.3% greater accuracy with this model than with the previous conventional SVM models.

PROPOSED SYSTEM

Our objective is to develop a heart disease prediction system to predict cardiac disease, this research uses a variety of machine learning techniques, such as the logistic regression, random forest, support vector machine, decision tree classifier, and KNN. The implementation language is called Python. The preprocessing of the dataset helps to improve accuracy by removing unnecessary data.

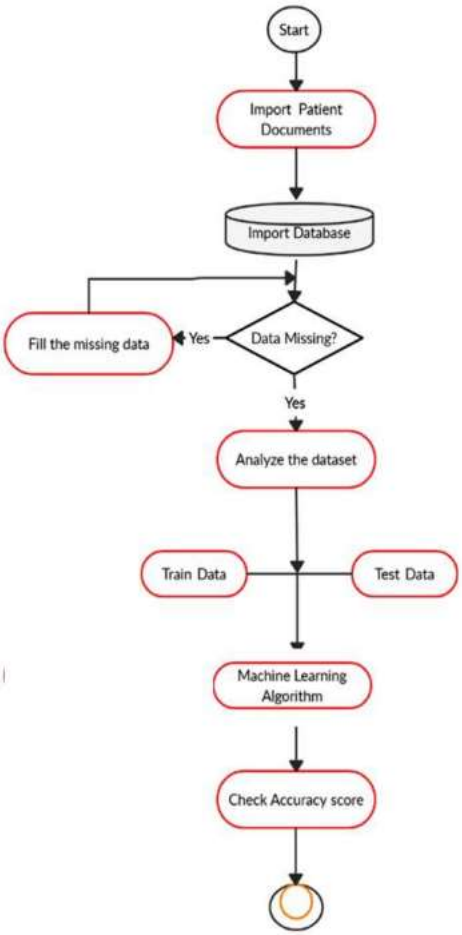
Dataset :- The Online repository's heart disease dataset was used for this investigation. The dataset with the following parameters: age, sex, type of chest discomfort, blood pressure at rest, serum cholesterol, etc. Following preprocessing, the dataset was divided into training 70% of the time and examination (30%). The models included the logistic regression, random forest, support vector machine, decision tree, and KNN are trained using the training data, and then the testing set was used for testing.

Advantages of Proposed System

Improved Efficiency :- The accuracy of cardiovascular risk prediction is greatly increased by machine learning techniques, allowing patients to be identified at an early stage of the disease and to benefit from preventive treatment.

Personalized Risk Assessment :- By taking into account each person's particular combination of risk factors, machine learning models can generate customized risk profiles. More specialized interventions and recommendations catered to an individual's unique health profile are made possible by this personalized approach.

Continuous Monitoring :- Over time, machine learning systems may make it possible to continuously monitor patient data. This makes it possible to dynamically modify risk assessments in response to evolving medical conditions and lifestyle choices, giving a more comprehensive and current image of a person's heart health.



FLOW DIAGRAM

METHODOLOGY

Our research is divided into two main sections: the first is the pre-processing stage, in which we select the most pertinent attributes; the second is the application of machine learning algorithms to determine which algorithm provides the best accuracy. Our The proposal has multiple phases, and a detailed explanation of the methodology can be found in

1. Dataset Collection :- In this instance, we make use of a dataset of individuals who have conducted tests and analyses to identify heart disease. The data set is a matrix, with the patients' rows denoting their characteristics and the factors or attributes (features) to be examined.

2.Manual Exploration :- In the creation of machine learning algorithms, this stage is crucial. Due to the fact that we rank or label each individual as sick or not after analysing the data set. We form the data set in order to provide the algorithms with the training dataset.

3. Data Pre-processing :- Because the quality of the data and the valuable information that can be extracted from it directly impact our model's capacity to learn, data pre-processing is a crucial stage in the machine learning process. Before putting our data into the model, preprocess it.

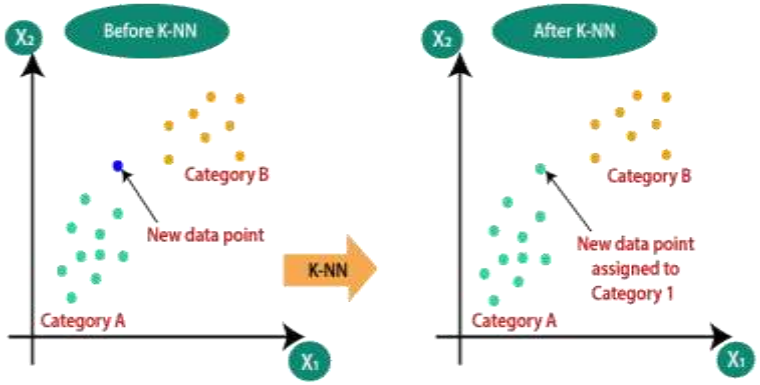
a)Features selection : Using the Pearson Correlation Method (Correlation matrix) [4], we can select the most significant features from a large set of features in the data set [2]. In order to apply machine learning algorithms and obtain better accuracy, we only select those attributes that are highly dependent on one another when attempting to detect links between them.

b) Splitting Dataset : After mentioning the target column in the data set, we split the data set into two smaller data sets in order to train the machine learning algorithm. Test-set is used to test the algorithm and Training-set is used to train it.

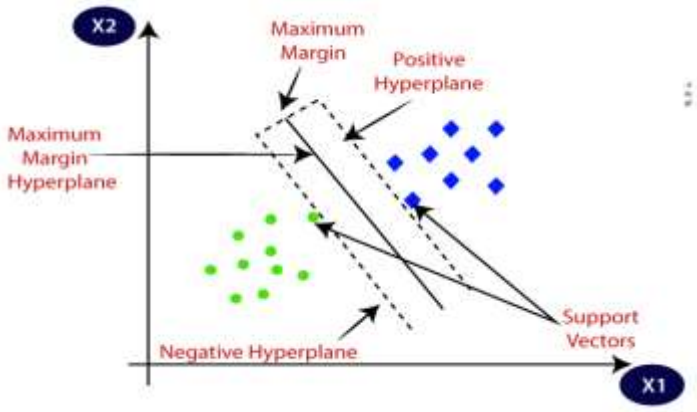
4. Modelling :- In order to determine which of the selected algorithms—Random Forest, Decision Tree, SVM, Logistic Regression, and KNN—is best, we need to apply and test them.

Algorithms application In this step, we discover that Random Forest, Decision Tree, SVM, Logistic Regression, and KNN are the most popular and effective algorithms. Our method relies on using the three algorithms on a variety of sized data sets to verify the accuracy of every one and, most importantly, identify the more stable one during training and undoubtedly produces excellent accuracy.

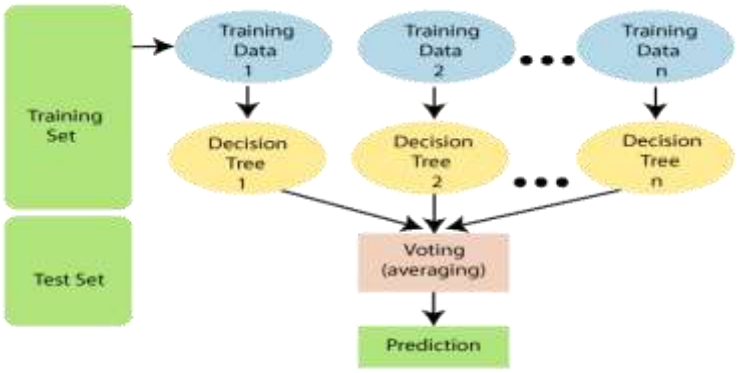
a)**KNN (K-Nearest Neighbors) :-** The K-NN algorithm ranks the k closest data points according to their distance from the input by utilizing soma metrics (such as Manhattan distance, Euclidean distance, etc.).Inputs are categorized based on the majority vote of adjacent categories. Although this model doesn't need to be trained, classification involves a lot of processing.



b) SVM (Suppot Vector Machine) :- It helps with problems involving both regression and classification. By drawing parallel lines to the hyperplane that passes through one of the closest points, it creates a marginal distance between them to ensure that they can be easily separated. Maximizing marginal distances is our goal. The points that genuinely cross the marginal plane that we made parallel to the hyperplane are known as support vectors. There might be more than one point.

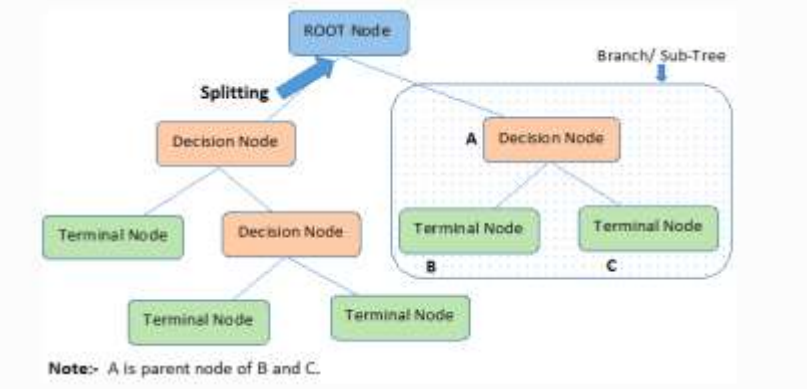


c) Random Forest :- Random forests and other supervised machine learning algorithms are frequently utilized for classification and regression issues. Regression analysis uses the mean of the majority votes for each classification and their respective majority votes. More trees in the forest will produce higher accuracy and ward off overfitting.



d) Decision Tree :- When determining whether a node in a decision tree should be divided into two or more sub-nodes, various algorithms are employed. A greater homogeneity is the outcome of the creation of sub-nodes. Alternatively, we can state that the purity of the node increases as the target value increases.

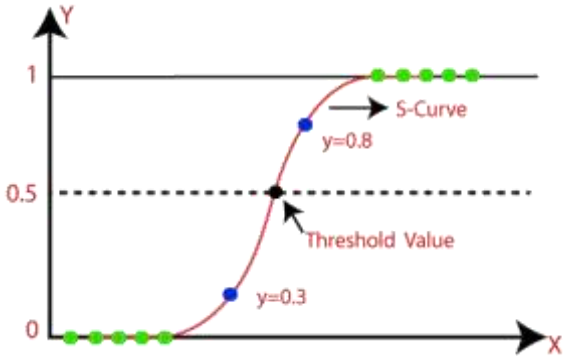
Internal nodes in a decision tree indicate the tests that have been conducted, branches indicate the decisions made based on those tests, and leaf nodes show the test results. The root node is the highest node. For our convenience, it breaks down large problems into smaller ones.



e) Logistic Regression :- One of the most popular categories of statistical models for predictive analysis and classification is the logit model. Logistic regression is a technique that calculates the probability of an event by utilizing a dataset of independent variables. Owing to the outcome's probability, the dependent variable is restricted to a range of 0 to 1.In logistic regression, the odds of success divided by the odds of failure are transformed using the logit transformation. The following formulas can also be used to represent a logistic function: This is also known as the natural logarithm of odds, or log odds.

$$Logit(pi) = 1/(1+ exp(-pi))$$

$$ln(pi/(1-pi)) = Beta\_0 + Beta\_1*X\_1 + \ldots + B\_k*K\_k$$



DESIGN

CONCLUSION

Heart conditions are a leading global cause of death. However, by assisting in the patients' improved state of health, their early diagnosis helps save lives. We've discussed a few of AI heart disease prediction systems. We examined many NL and AI algorithms in an attempt to compare and identify which has the best features. Each algorithm has generated a unique result under a range of conditions. Subsequent investigation reveals that the heart disease prediction model only has a very low level of accuracy; hence, more complex models are needed to increase accuracy and provide early heart disease predictions. Future approaches for simple, inexpensive, and highly accurate early detection of heart disease will be suggested.

Machine learning is crucial to the study of disease prediction. This work forecasts cardiac illness using a variety of machine learning techniques. The outcomes of the experiment demonstrate that the Random Forest algorithm attains the maximum accuracy of 91.8%, effectively fulfilling the goal of enhancing prediction accuracy. Future research will focus on delving deeper into evolutionary computation methods for the given challenge and evaluating their efficacy.

Heart disease is a serious problem in the world that is expanding right now. Thus, an automated system to forecast heart disease at an earlier stage is required. In order to help the doctor diagnose patients more quickly, as well as to benefit the public, who can use this automated system to keep track of their health problems. This paper summarised some of the expert automated systems. Prediction and feature selection are two fundamental components of any automated system. We can improve our ability to predict heart disease by making effective feature choices. We have provided a summary of a few algorithms, such as the random search algorithm and the hybrid grid search algorithm, that are helpful when choosing the features. Therefore, it will be preferable to use search algorithms in the future to choose the features, and then machine learning techniques will be applied to predict heart disease with better outcomes.

RESULT

This paper's main goal is to increase the accuracy of the heart disease rate forecast. Five methodologies, namely Logistic Regression, Random Forest, SVM, Decision Tree Classifier, and KNN, were employed to conduct simulation-based experiments. The results show that, when compared to the other four techniques, the random forest method provides higher accuracy. A training set and a testing set are created by splitting the used data set. In this case, training uses 70% of the dataset, with the remaining portion going towards testing. The dataset indicates that a higher number of individuals with heart disease fall within the 50-60 age range.

FUTURE SCOPE

Future updates to this application could include the ability to notify all of the user's family members in advance if he develops heart disease. Additionally, the closest hospital should be informed of the information. Online medical consultations with the closest physician should be another feature. It is also possible to suggest new algorithms to increase accuracy and dependability. Large volumes of user data from all over the world can be stored using big data technologies like Hadoop, and user data and reports can be managed by utilising technologies like cloud computing. It is noteworthy that machine learning (ML) applications, utilising a range of effective algorithms, are employed not only in the diagnosis and prognosis of diseases, but also in the fields of radiology, bioinformatics, and medical imaging diagnosis, among others.

REFERENCE

[1] Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259-64). IEEE. ICCRDA 2020 IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012072 IOP Publishing doi:10.1088/1757-899X/1022/1/012072 9

[2] Bhunia, P.K.; Debnath, A.; Mondal, P.; D E, M.; Ganguly, K.; Rakshit, P. Heart Disease Prediction using Machine Learning. *Int. J. Eng. Res. Technol.* **2021**, *9*.

[3] Shah, D.; Patel, S.; Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. *SN Comput. Sci.* **2020**, *1*, 345.

[4] Khan, I.H.; Mondal, M.R.H. Data-Driven Diagnosis of Heart Disease. *Int. J. Comput. Appl.* **2020**, *176*, 46–54

[5] Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 261–268.