IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

YERRA RENU SREE, Prof. M. RAMJEE

Department of Computer Science and Technology

College of Engineering (A), Andhra University, Visakhapatnam, AP.

Abstract

Machine Learning is finding applications in a wide number of disciplines throughout the world, including the healthcare industry. It can be used to predict the existence or lack of locomotor disorders, cardiac illnesses, and other conditions. When these assumptions are established well in advance, it means they can provide vital insights to clinicians, allowing them to personalise diagnoses and therapies to each individual patient. The goal of this study is to apply algorithms based on machine learning to anticipate possible cardiac problems in individuals. A comparative investigation of classifiers such as K-Nearest Neighbours (K-NN), Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Random Forest is part of the project. It also presents an ensemble classifier which employs both strong as well as weak classifiers to conduct hybrid classification. This method is helpful since it allows for the incorporation of a wide range of validation and training samples, which leads to improved precision and predictive analysis. This Model accepts as input the qualities that are kept after the feature selection techniques have been applied. Only 14 of the 75 features that make up the UCI Heart Disease data set are chosen and utilised for prediction. This model predicts whether a person has heart disease or not.

Keywords: heart diseases, Machine learning, comparative analysis of classifiers and python.

Introduction

"Machine Learning is a method of manipulating and extracting implicit, previously unknown/known, and potentially useful data information" [1]. Machine Learning constitutes a huge and complex science, and its scope and application are expanding all the time. Machine learning incorporates a variety of classifiers from Supervised, Unsupervised, and Ensemble Learning that are used to predict and determine the accuracy of a given dataset. We could apply that information to our HDPS initiative, which will benefit many individuals. Cardiovascular illnesses are a term that describes a variety of problems that can damage your heart. According to the World Health Organisation, there are 17.9 million world-wide deaths at (Cardiovascular diseases) CVDs [2].

It is the leading cause of mortality in adults. Our study can identify who is likely to be diagnosed with heart disease based on their past medical conditions [3]. It recognises who has any symptoms of heart illness, such as chest discomfort or high blood pressure, and can assist in identifying disease with fewer medical tests and effective therapies, allowing them to be healed accordingly.

This project is primarily concerned with three data mining techniques: (1) logistic regression, (2) KNN, and (3) Random Forest Classifier. Our project's accuracy is 87.5%, which is higher than the prior system, which employed merely a single data mining approach. As a result, employing additional data mining techniques improved HDPS precision as well as effectiveness. Logistic regression is an example of supervised learning. In logistic regression, only discrete values are employed.

A peaceful This work was prompted by a significant amount of work linked to the detection from Cardiovascular Heart Disease utilising Machine Learning techniques. This paper includes a quick review of the literature. Various algorithms, such as Random Forest Classifier, Logistic Regression, KNN, and others, were used to predict Cardiovascular Disease. The results show each and every algorithm has a different ability to register the given objectives [4].

The model using IHDPS may determine the decision boundary utilising the old and new machine learning as well as deep learning models. It facilitated the most significant and fundamental factors/knowledge, including family history of any cardiac disease. However, the accuracy gained in such IHDPS models was significantly less than that of the new impending model for identifying coronary heart disease utilising artificial neural networks and other machine and deep learning methods. McPherson et al.,[5] identified the risk variables associated with coronary coronary artery disease or atherosclerosis using an internal implementation algorithm that employs certain Neural Network methods and was only able to effectively predict whether the test subject was suffering from the specified ailment or not.

R. Subramanian et al. [10] proposed employing neural networks to diagnose and predict heart disease, blood pressure, and other features. A deep Neural Network had been constructed that incorporated the given disease attributes and was able to produce an output that was determined by the output perceptron as well as nearly included 120 hidden layers, that is the basic and most relevant technique for ensuring a precise identification regarding heart disease if the model is used for Test Dataset. The supervised network is being recommended for heart disease diagnosis [6].

When a doctor tested the model using unknown data, the model utilised and trained from previously learned data and projected the result, determining the precision of the supplied model. The identification of cardiac disease is the most difficult problem. There are tools that can forecast heart disease, but they are either too expensive or too inefficient to quantify the risk of cardiovascular disease in humans. Early identification of heart disorders has been shown to reduce mortality and overall consequences. But it is not feasible to correctly monitor patients on a daily basis in all circumstances, and consultation with a doctor for 24 hours is not accessible since it needs more intelligence, time, and skill. We are able to employ different algorithms for machine learning to analyse information to identify hidden patterns in today's environment since we have a lot of data. Hidden patterns in medical data can be utilised for health diagnosis.

Existing System:

Heart disease has the distinction of being dubbed a "silent killer," as it may kill a person without causing noticeable symptoms. The illness's nature is the source of increased concern about the sickness and its effects. As a result, ongoing attempts are being made to forecast the probability of this devastating disease in advance. As a result, numerous technologies and approaches are being tested on a regular basis to meet today's health demands. Machine Learning methods can be quite useful in this area. While heart disease can manifest itself in various ways, there is an established set of key risk factors that impact whether someone is at danger of heart illness or not. We can draw conclusions by gathering data from numerous sources, categorising it under appropriate headings, and lastly evaluating it to get the needed facts. This approach can be extremely successfully adapted to undertake heart disease prediction. As the well-known adage goes, "prevention is better than cure," and early prediction and control of heart disease can assist to avoid and reduce mortality rates from heart disease.

Proposed System:

The system's operation begins with the gathering of data and the selection of the key properties. The necessary data is then preprocessed into the suitable format. The data is subsequently divided into data for testing and training. Using the training data, the algorithms are put into effect as well as the model is trained. The system's correctness is determined by testing it employing the testing data.

Dataset

An organised dataset of individuals was chosen with their history of cardiac issues and other medical diseases in mind. Heart illness refers to the various disorders that affect the heart. Cardiovascular disorders are the leading cause of mortality among middle-aged persons, according to the World Health Organisation (WHO). We use a data set that contains the medical histories of 304 distinct patients of various ages. This dataset provides us with much-needed information, such as the age of the individual, fasting sugar level, resting blood pressure, and so on, which aids us in determining if the patient has a cardiac condition or not. This dataset comprises 76 medical features of various patients that enable us determine whether the patient has a risk of acquiring a heart condition or not, as well as classify individuals who are at danger and those who are not. This dataset on heart disease is obtained from the UCI source. This dataset extracts the pattern that leads to the discovery of patients at risk of developing heart disease.

S.NO	Attribute	Description	Type
1	Age	Patient’s age (29 to 77)	Numerical
2	Sex	Gender of patient (male-0 female-1)	Nominal
3	CP	Chest pain type	Nominal
4	Trest bps	Resting blood pressure (in mm Hg on Admission to hospital, values from 94 to 200)	Numerical
5	Chol	Serumcholesterolinmg/dl, values from126 to 564)	Numerical
6	FBS	Fasting blood sugar>120mg/dl, true- 1 false-0)	Nominal
7	Resting	Resting electro cardio graphic result (0 to1)	Nominal
8	Thali	Maximum heart rate achieved (71 to 202)	Numerical
9	Exang	Exercise included angina (1-yes 0-no)	Nominal
10	Old peak	ST depression introduced by exercise Relative to rest (0 to .2)	Numerical
11	Slope	The slope of the peak exercise ST segment (0 to1)	Nominal
12	Ca	Number of major vessels (0-3)	Numerical
13	Thal	3-normal	Nominal
14	Targets	Presence or Absence	Nominal

Table 1: Dataset Description

Methodology

This study analyses different machine learning methods, including K closest neighbours (KNN), Logistic Regression, and Random Forest Classifiers, which can assist clinicians or medical analysts in properly diagnosing Heart Disease. This documentation involves reviewing journals, published papers, and recent data on cardiovascular illness. Methodology provides a structure for the suggested model [7]. The technique is a procedure that involves stages that convert provided data into recognised data patterns to facilitate consumers' understanding.

The suggested approach consists of three stages, the first of which is data collection, the second of which is substantial value extraction, and the third of which is data exploration.Depending on the procedures employed, data preparation addresses missing values, data cleansing, and data normalisation. Following data pre-processing, classifiers are used to categorise the pre-processed data. The classifiers utilised in the model that is suggested are DecisionTree, svm, KNN, Logistic Regression, and Random Forest Classifier.

Finally, the suggested model is implemented, and we assessed our model for accuracy and performance using several performance indicators. Using many classifiers, this model has created an excellent Heart Disease Prediction System. This model predicts using 13 medical characteristics such as chest pain, cholesterol levels, fasting sugar, blood pressure, age, gender, and so on [8].

We can see from these results that, while most studies use various algorithms such as SVC, Decision tree for the identification of patients with a diagnosis of coronary artery disease, Random Forest Classifier, KNN, and Logistic regression outperform them [9]. The algorithms that have been utilised are more accurate, save a lot of money, and are faster than the methods that earlier researchers used. Furthermore, the greatest accuracy attained by KNN is equivalent to 100%, making it greater or nearly identical to the accuracies obtained in prior studies. So, to summarise, our accuracy has improved as a result of the added medical characteristics included in the dataset we utilised.

Our research also shows that KNN surpasses Random Forest Classifiers in the prediction of patients with heart disease. This demonstrates that KNN is superior in the diagnosis of cardiac disease.

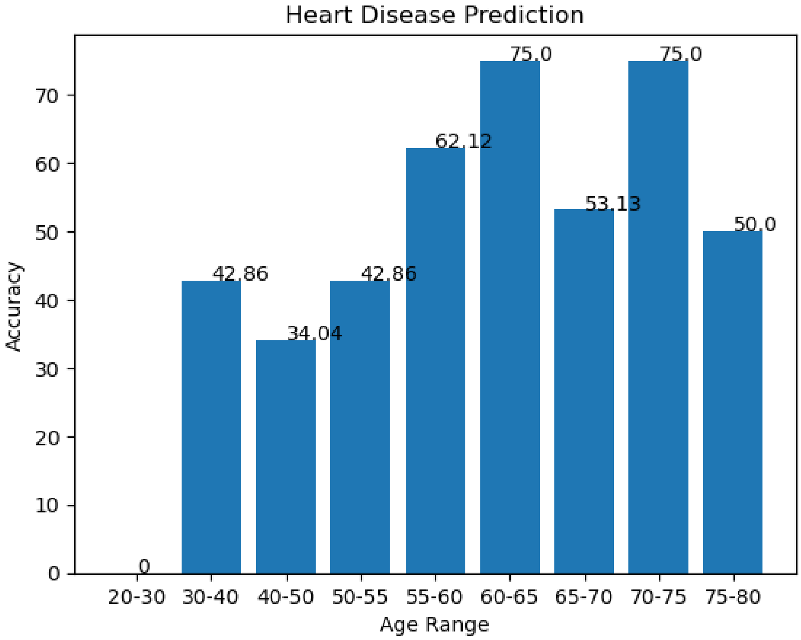


Fig 1: Heart disease prediction on the basis of their age.

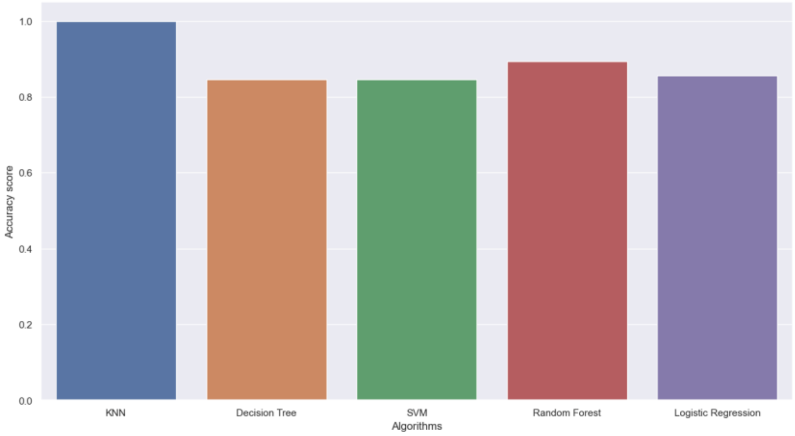


Fig 2: Accuracy for different classifiers.

Using the PYTHON code, a background image is created to easily present the graphical representation of candidate’s heart diseases. In this program, once the candidate’s ID number is entered, softwre can predict the person’s age, BP, cholesterol etc.

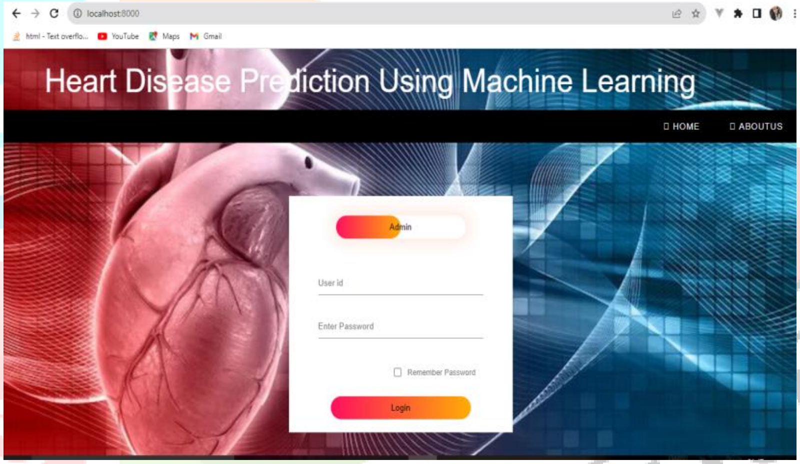
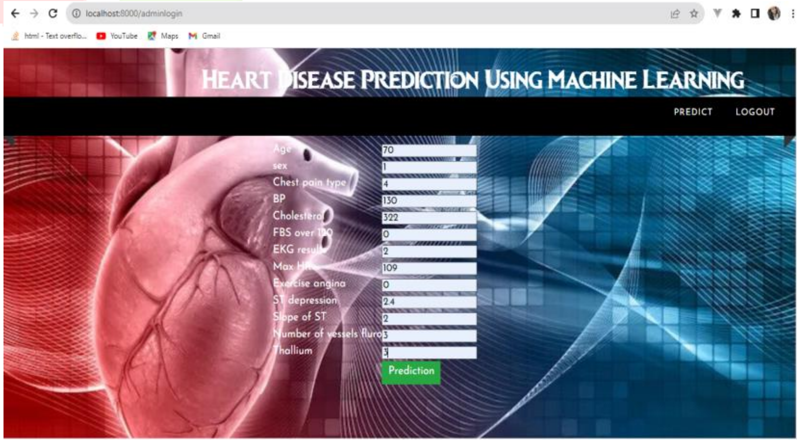


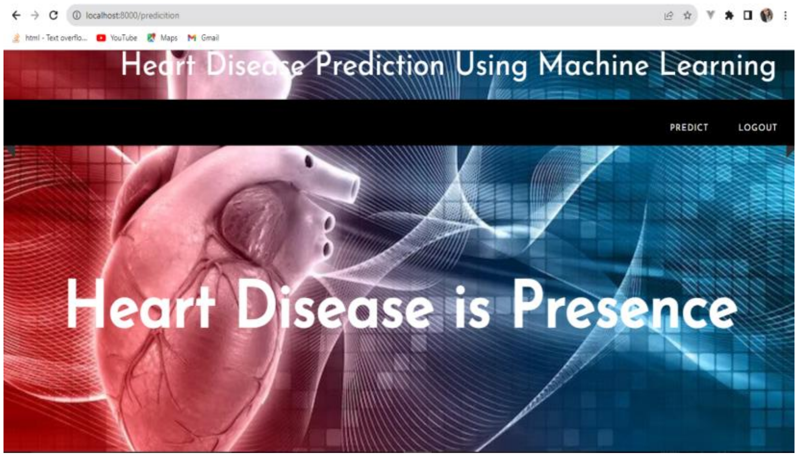
Fig 3: Final output for testing candidate.


localhost:8000/adminlogin
html - Text overflo...
YouTube
Maps
Gmail
HEART DISEASE PREDICTION USING MACHINE LEARNING
PREDICT
LOGOUT
Prediction

Age 70
sex 1
Chest pain type 4
BP 130
Cholesterol 322
FBS over 120 0
EKG results 2
Max HR 109
Exercise angina 0
ST depression 2.4
Slope of ST 2
Number of vessels fluro 3
Thallium 3

Fig 4: Prediction information for candidate.

The above image shows the candidate’s Age, Sex, Cp, Trestbps, Chol, Fbs, Resting, Thali, Exang, Old peak, Slope, Ca, Thal and Targets. As per the detail information and values, the software can predict the presence or absence of heart diseases.



Heart Disease Prediction Using Machine Learning

Heart Disease is Presence

Fig 5: Final window for prediction.

Conclusion

Heart disease is an important threat in India as well as around the world; using promising technology such as machine learning to detect the earliest signs of heart disease would have a significant influence on society. Early detection of cardiac disease can help high-risk individuals make lifestyle adjustments and, as a result, prevent consequences, this can be a significant step forward in the area of medicine. Every year, more people are diagnosed with heart disease. This necessitates an early diagnosis and therapy. The use of appropriate technological assistance in this area may be extremely valuable to the medical community and patients. In this work, the seven distinct machine learning methods utilised to test performance on the dataset include Random Forest, SVM, Decision Tree, and Logistic Regression.

The predicted qualities contributing to heart diseases in patients are present in the dataset, which comprises 76 variables, 14 of which are crucial for evaluating the system. If all of the features are taken into account, the author's system has a lower efficiency. Feature selection is done to improve efficiency. In this case, n features have to be chosen in order to evaluate the model with more precision. Because the correlation of several characteristics in the dataset is nearly identical, they are discarded. When all of the attributes within the dataset are included, the efficiency drops dramatically.

The accuracies of all five machine learning approaches are evaluated, and one prediction model has been created as a result. As a result, the goal is to employ several assessment criteria, such as the confusion matrix and accuracy, to efficiently anticipate the disease. When all five are compared, the KNN has the greatest accuracy of 100%.

References

[1] Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-8

[2] Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

[3] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.

[4] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[5] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.

[6] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299-303). IEEE.

[7] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. IEEE antennas and propagation magazine, 58(5), 84-92.

[8] Buechler K F & McPherson P H (1999). U.S. Patent No. 5,947,124. Washington, DC: U.S. Patent and Trademark Office.

[9] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. International journal of epidemiology, 18(2), 361-7.

[10] Kiyasu J Y (1982). U.S. Patent No. 4,338,396. Washington, DC: U.S. Patent and Trademark Office.