



Heart Disease Prediction using Machine Learning

¹Sibhuraj Pappu, ²Hamza Girach, ³Mohd Zaid Gandhi, ⁴Adnan Hathiyari, ⁵Mohammed Juned

^{1,5} Assistant Professor

²⁻⁴Students, ¹⁻⁵Department of Computer Engineering,

¹⁻⁵Rizvi College of Engineering Mumbai, India

Abstract: Heart disease is considered one of the leading causes of death worldwide. Prediction is a difficult task that requires specialized and advanced knowledge and cannot be easily predicted by a doctor. Healthcare is still "rich in information" but "poor in knowledge". Within the medical system, a lot of data is available on the Internet. However, there is a lack of effective analytical tools to reveal hidden relationships and patterns in the data. Medical diagnostic automation systems increase medical efficiency and reduce costs. This web application aims to predict the outbreak of illness based on data obtained from medical research conducted by the Kaggle and Cleveland Foundation, especially in the field of heart disease. The purpose is to extract hidden patterns by applying data mining techniques to datasets. It is noteworthy for heart disease and predicts the presence of heart disease in patients whose presence is assessed on a scale. Predicting heart disease requires large amounts of data that are too complex to process and analyze using traditional techniques.

Key Words: Prediction, heart disease, machine learning, algorithms, analysis.

1. INTRODUCTION

The maximum mortality of each India and overseas is specifically due to coronary heart sickness. According to World Health Organization (WHO), coronary heart associated illnesses are liable for the taking 17.7 million lives each year, 31% of all worldwide deaths [6]. Hence, that is important time to test this demise fee through figuring out the sickness efficaciously withinside the preliminary stage. We can use records mining technology to find out expertise from the datasets. The located expertise may be utilized by the healthcare directors to enhance the nice of carrier. The located expertise also can be utilized by clinical practitioners to lessen the range of detrimental drug effect, to indicate much less steeply-priced therapeutically equal alternatives. Anticipating affected person's destiny behavior at the given records is one of the essential programs of records mining strategies that may be utilized in healthcare control.

A most important undertaking going through healthcare corporations (hospitals, clinical centres) is the supply of nice offerings at low-priced costs. Quality carrier implies diagnosing sufferers efficaciously and administering remedies which might be effective. Poor scientific selections can result in disastrous outcomes which can be consequently unacceptable. Hospitals ought to additionally decrease the value of scientific tests. They can obtain those effects through using suitable computer-primarily based totally facts and/or choice guide systems. Healthcare records is massive[8]. It consists of affected person records, useful resource control records, and converted records. Healthcare corporations ought to have the capacity to examine records. Treatment statistics of tens of thousands and thousands of sufferers may be stored, and automated and records mining strategies may also assist in answering numerous essential and important questions associated with fitness care. This suggestion is promising as data modelling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

2. LITERATURE SURVEY

Intelligent Heart Disease Prediction System Using Data Mining Technology: The medical industry has accumulated vast amounts of health data, but unfortunately to reveal hidden information for effective decision making. Not "mined". The discovery of hidden patterns and relationships is often undeveloped. Advanced data mining technology is useful. This study prototyped an intelligent heart disease prediction system. It uses data mining techniques: decision trees, naive Bayes, and neural networks.

Smartphone-based ischemic heart disease (heart attack) risk prediction: Android-based prototype software was developed by integrating clinical data from patients admitted with IHD (ischemic heart disease). Analyzing clinical data from 787 patients, it may show hypertension, diabetes, dyslipidemia (abnormal cholesterol levels), smoking, family history, obesity, stress, and potentially unrecognized IHD. Correlated with risk factors such as existing clinical symptoms. The data is mined using data mining technology and scores are generated. The risk of IHD is divided into low, medium, and high.

Analysis of data mining techniques to predict heart disease: heart disease is considered one of the leading causes of death worldwide. Prediction is difficult for medical professionals as it is a difficult job that requires higher knowledge and expertise to predict. This article deals with the problem of predicting heart disease based on input attributes based on data mining techniques. We studied heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques was measured by combining the results of the prediction accuracy, ROC curve, and AUC value using the standard 6 dataset as well as the collected data set. ten. Based on the performance factor, the SMO and Bayes Net techniques showed optimal performance compared to the performance of the KStar, Multilayer Perceptron and J48 techniques.

Machine Learning Application of Coronary Atherosclerosis Risk Prediction: Coronary artery disease is the leading cause of death worldwide. In this study, we propose an algorithm based on machine learning techniques to predict the risk of coronary atherosclerosis. The peak expectation maximization (REMI) technique is proposed to estimate the missing values in the atherosclerotic database. The conditional probability maximization method is used to remove irrelevant attributes and reduce the size of the object space and thus improve the learning rate. STULONG and UCI databases are used to evaluate the proposed algorithm. The performance of heart disease prediction for two classification models was analyzed and compared with previous work. The test results show the percentage of improved accuracy of risk prediction of our proposed method compared with other works. The effect of the addition of missing values on the prediction performance was also evaluated and the proposed REMI approach is significantly more efficient than conventional techniques [4].

3. OBJECTIVES

- Easy to use:

The fundamental goal of this assignment is to expand a platform as a way to be easy and smooth to use, as right here one should offer the affected person's scientific info and primarily based totally at the capabilities extracted the set of rules will then locate the coronary heart disorder and notice its kind. As right here set of rules does the challenge therefore a well- educated version is much less sure to make mistakes in predicting the coronary heart disorder and its kind therefore, in brief accuracy is advanced and thereby it additionally saves time and makes simpler for medical doctors in addition to sufferers to are expecting whether or not they may be vulnerable to any kind of coronary heart disorder or not, that's in any other case we hard to do without doctor's involvement.

- No human intervention required:

To locate the coronary heart disorder one should offer scientific info along with age, cholesterol, etc. and right here the set of rules will offer the consequences primarily based totally at the capabilities extracted and therefore right here probabilities of mistakes been made are very minimal seeing that there may be no human intervention and it additionally saves lot of time for the sufferers or medical doctors and they could similarly continue for remedies or different approaches should quicker. This is in case whilst consequences are furnished quicker to them. This can in-flip make the precaution/prevention technique of coronary heart remedy loads quicker whilst it saves medical doctors and affected person the important time, so that it will pass directly to similarly remedies and precautions to be taken to decrease the effect of that coronary heart disorder.

- Detect the coronary heart disorder kind:

In this assignment our goal isn't simplest to locate and are expecting the kind of coronary heart disorder however pin factor closer to the effect of the coronary heart disorder. Getting hints on precautions to be taken will assist the medical doctors and sufferers to development effortlessly to similarly steps of their remedy.

- Efficient use of to be had annotated records samples:

There is huge consent that a success education of gadget gaining knowledge of algorithms calls for many thousand annotated education samples. Hence, we use a community and education strategy.

4. PROPOSED SYSTEM

Prediction of heart disease is a web-based web learning application, formed by a UCI data set. Users enter their specific medical information to get predictions about heart disease for that user. The algorithm will calculate the probability of the presence of heart disease. The result will be displayed on the site itself. Therefore, minimize the cost and time needed to predict the disease. Data format plays an important part in this application. At the time of download, the user data application will check its appropriate file format and if the dialog box should not be invited. Our system will perform the following algorithm:

- Logistic regression.

The algorithms will be trained using the data set obtained from Kaggle.75% items of the data set will be used to train and the remaining 25% to check the accuracy of the algorithm. In addition, some steps will be done to optimize algorithms, thus improving accuracy. These steps include cleaning data sets and pre-processing data. The algorithms are found on their precision and people observed that SVM is the most accurate in three with 64.4% efficiency. Therefore, it is selected for the main application. The main application is a web application that accepts different user settings as input and calculating results. The result is displayed with the precision of anticipation.

1) Website:

The system will include a website where users are registered to get the health rate of their hearts about predictive analysis of their heart disease. Users will have to fill in an original form to register. After that, the user will be redirected to

the profile page where they will have to fill out their records by completing all information about their hearts. After sending health information, the patient will be able to check the report where they will know the status or risk of their hearts

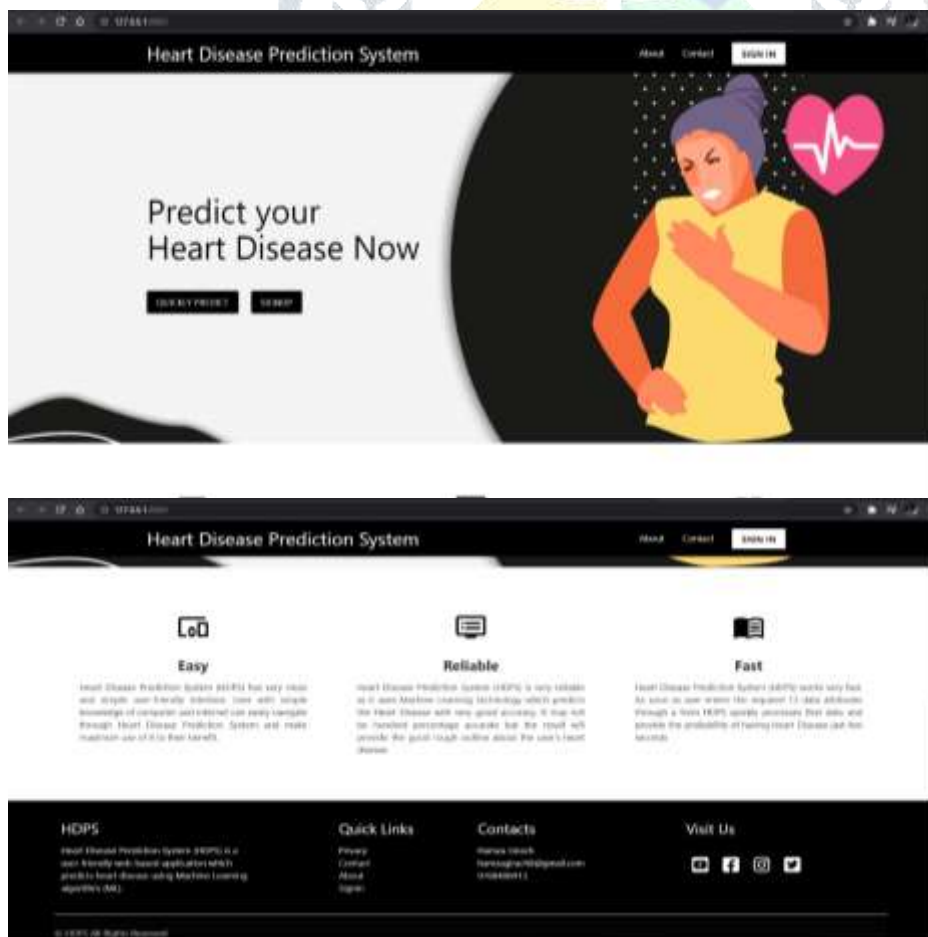
2) Machine Learning algorithm:

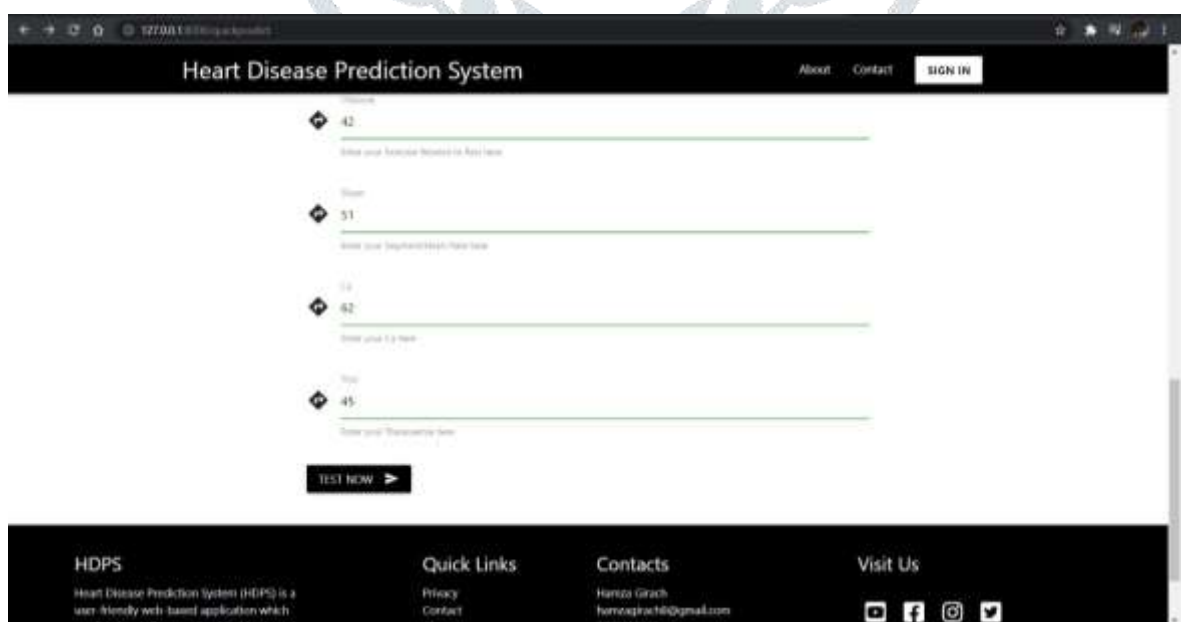
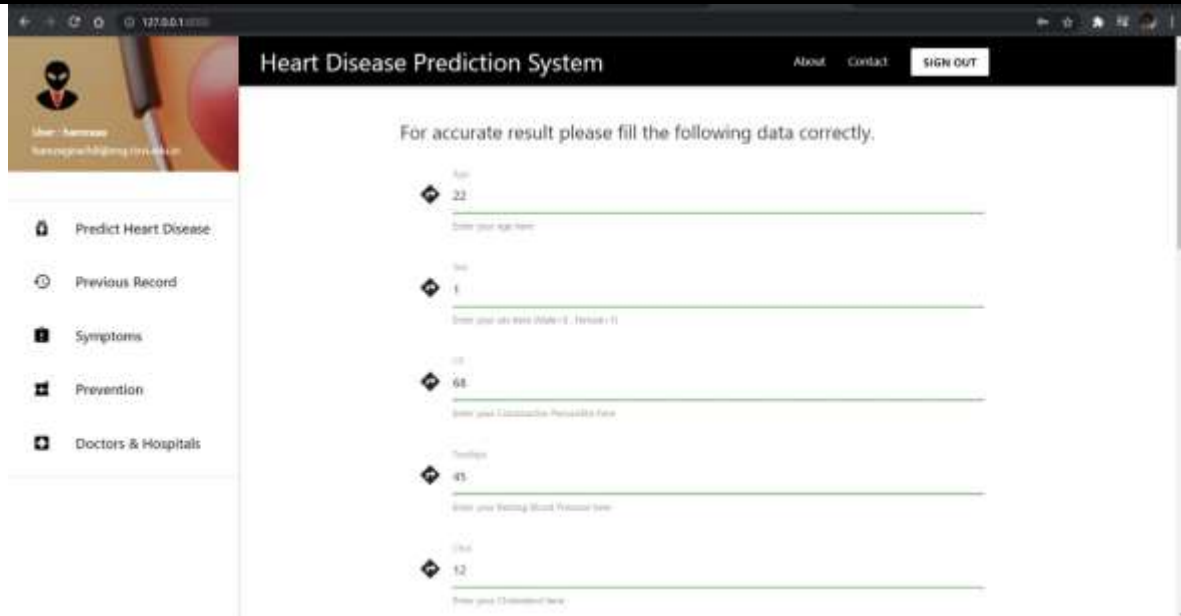
Machine Learning algorithms will be used to predict the risk of heart disease in percentage.

5. METHODOLOGY

- **Data pre-processing:**
The data set we obtained is not completely accurate and free of errors. So first we will do the following on it.
- **Data cleaning:**
The NA values in the dataset are the main drawback for us as it will significantly reduce the accuracy of the prediction. Therefore, we will remove nullable fields. We'll replace it with the column mean. This way we will remove all values from the dataset.
- **Feature scaling:**
Because the range of raw data values varies widely, in some machine learning algorithms the target functions will not perform well without feature scaling.
- **Factorization:**
In this section, we have assigned meaning to the values so that the algorithm does not confuse them. For example, assign a meaning to 0 and 1 in the age section so that the algorithm does not consider 1 to be greater than 0 in this section.
- **Supported Vector Machine:**
Supported Vector Machine (SVM) is a supervised learning method for data analysis used for classification and regression analysis. It is given a training dataset, marked as belonging to one of two categories, then the SVM training algorithm builds a model that assigns new examples to one type or the other, making it become a non-probability binary linear classifier. The SVM model is a representation of the examples as points in space, mapped to examples of are mapped in that same space and are predicted to fall into the category based on which side of the distance they are on.

5. RESULT





Heart Disease Prediction System

About Contact SIGN OUT

User: Admin
hanuzgtrah@gmail.com

Predict Heart Disease
Previous Record
Symptoms
Prevention
Doctors & Hospitals

Your probability of having Heart Disease is 0.2861140458117038

Your chance of having Heart Disease is low.
Please continue your healthy habit.

HDPS Quick Links Contacts Visit Us

Heart Disease Prediction System

About Contact SIGN OUT

User: Admin
hanuzgtrah@gmail.com

Predict Heart Disease
Previous Record
Symptoms
Prevention
Doctors & Hospitals

Here is the record of your previous Prediction of Heart Disease.

VIEW DETAILS

S.No	Date	Probability
1	April 21, 2022	0.2861140458117038

HDPS
Heart Disease Prediction System (HDPS) is a user-friendly web-based application which predicts heart disease using Machine Learning algorithm (ML).

Quick Links
Privacy
Contact
About
Signin

Contacts
Hanuz Gtrah
hanuzgtrah@gmail.com
9750406413

Visit Us
YouTube Facebook Instagram Twitter

© HDPS All Rights Reserved

6. CONCLUSION

This paper conclude that the model is develop as a working model which can predict whether a person has any heart disease by looking at different parameters, and determine whether the person has any existing threat to any kind of pulmonary disease. Thus this paper represents an overview of Heart Disease Prediction using Machine Learning.

7. REFERENCE

- [1] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, July 2015
- [2] M. Raihan, Saikat Mondal, Arun More, Md. Omar Faruq Sagor, Gopal Sikder, Mahub Arab Majumder, Mohammad Abdullah Al Manjur and Kushal Ghosh "Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design", September 2014.
- [3] Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin "Analysis of Data Mining Techniques for Heart Disease Prediction", May 2015.
- [4] Soodeh Nikan, Femida Gwady-Sridhar, and Michael Bauer "Machine Learning Application to Predict the Risk of Coronary Artery Atherosclerosis", IEEE, August 2016
- [5] Sanjay Kumar Sen Asst. Professor, Computer Science & Engg. Orissa Engineering College, Bhubaneswar, Odisha – India." Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms" International Journal of Engineering and Computer Science. Volume 6 Issue 6, June 2017
- [6] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja "Heart disease prediction using machine learning tech : A survey" International Journal of Engineering & Technology, 7 (2.8), April 2018.
- [7] Heart Disease Dataset - <https://www.kaggle.com/c/heart-disease> dated: Sept 2018
- [8] K. Srinivas, B. Kavihta Rani, A. Govrdhan "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attack" IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010.