IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE PREDICTION SYSTEM USING LOGISTIC REGRESSION ALGORITHM

[1]Nemata Ahamada Hassane, [2]V. Vani, [3]K. Ezhilarasan, [4]M. Narmatha

[1&2&3]PG Students, [2]Assistant Professor,

[1 &2&3&4]Department of Computer Science,

[1&2&3&4]Sri Krishna Arts and Science College, Coimbatore, India

Abstract: The heart is the main organ of a human being. Heart disease is increasing and expanding in the world. Many people are died due to cardiovascular disease. Detecting the disease at a premature stage may save the life of the patient. Data mining techniques are very popular and have been used in many fields including healthcare to help the doctor to make better decisions. Machine learning provides classification algorithms such as decision tree (DT), Naïve Bayes algorithm, Support machine vector (SVM), and Logistic Regression (LG) are used in many types of research for predicting heart disease. The dataset is collected from the Kaggle repository. It contains 604 data and 14 attributes used to train the model that will be used in the web application. Building an efficient prediction model to be deployed into the web application is the main objective of this project. A heart disease prediction model built using a logistic regression algorithm with an accuracy of 62% is the best one compare to others.

KEYWORDS **- heart disease, data mining, machine learning, classification, naïve Bayes, support vector machine, decision tree, logistic regression.**

I. INTRODUCTION

Heart disease is a harmful disease that affects the functionality of the heart. Heart disease represents a collection of different diseases such as heart failure, coronary artery disease (CAD), Heart Arrhythmia, Heart Valve Disease, Pericardial Disease, Cardiomyopathy (Heart Muscle Disease), and congenital heart disease that many people suffering with. Cardiovascular disease is one of the most dangerous diseases of the heart. As per the world health organization report, there are 17.9 million deaths occur through worldwide due to cardiovascular disease. Heart issue leads to scarier people's lifestyle. Now a day's cardiovascular diseases are very common. Cardiovascular disease describes a range of conditions that affect the heart mechanism. So cardiovascular disease is identified and measured by taking a range of conditions from the human affected heart. This project can predict and detect diagnose with heart disease from their medical history reports. It can help those who are having heart disease symptoms like high blood pressure, asthma, heart valve pain, and chest pain by giving effective treatment, accurate with less medical testing. So that patient can be cured with surety. It can reduce the death rate.

This HD (heart disease) prediction uses ML (Machine Learning) developed by regression algorithm. HD Prediction built by Logistic regression algorithm in the rate of 62%. The regression algorithm falls under the category of machine learning technique. The regression algorithm is widely used in skin cancer prediction techniques and breast cancer prediction. Improper diet, sugar at a young age, increasing age, taking more calories food and no physical activities are the main reasons for the heart disease. These are impacting major heart disease. Meditation, physical activities, a proper diet, and healthy food can protect from heart disease.

The objective of the HD prediction is to detect the heart disease from their age, name, medical history report, cardiovascular test, Blood test, Electrocardiogram, nuclear cardiac stress test, and so on. A dataset is kept from the Kaggle repository with the use of attributes and patient medical history report. Using a dataset and 14 types of attributes can predict and detect heart disease. In order to predict disease as early as possible, it can cure heart disease accordingly. The attributes are checked with help of a regression algorithm. Finally, this method can diagnose and predict patient has heart disease or not. It can provide Fewer medical tests with effective treatments with low-cost efficiency.

II. RELATED WORK

The diagnosis of cardiovascular heart disease using machine learning has motivated this project. This project has a survey of diagnosed heart disease. A quite significant amount of work helped to achieve the project goal using machine learning. The high-level efficient algorithms are in machine learning to detect and cure cardiovascular heart disease. This project was developed by a logistic regression algorithm using machine learning and each and every efficient algorithm can perform individual results for users analyzing and predicting the heart disease.

This system helps to calculate the decision boundary from the patient medical history report. This project gives common knowledge about the family history connected with cardiovascular heart disease but the accuracy of HDPS is less than other upcoming models such as coronary heart disease using other machine learning and artificial neural network. The most difficult is identified by McPherson et al using the algorithm of artificial neural networks [7]. Artificial neural networks detect whether the patient suffering from coronary heart disease or not and this can predict the heart disease. Symptoms of heart disease such as high blood pressure, and chest pain diagnosed using the artificial neural network which is introduced by R. Subramanian et al.

The deep artificial neural network was developed by given attributes and included 120 layers of attributes to provide accuracy-test outputs. Users can be diagnosed with which level of condition they are in and detect the accuracy type of heart disease using an accuracy test report. When this model was done by doctors by giving their unfamiliar data, this model produces the accuracy data calculated from previous health test reports. So, the doctor can provide the accurate type of heart disease and can give treatment accordingly.

III. PROPOSED ARCHITECTURE

Machine learning algorithms become very popular and used in different fields such as healthcare, business, etc. to solve many problems.in this system, we proposed a logistic regression machine learning algorithm for predicting heart disease. The logistic regression algorithm shows a high accurate result for the prediction. The user-friendly web application is developed using flask, HTML, and CSS. The user will login to the system and gives the required input for prediction.

3.1 Advantages

- Less time consuming and quick result
- Accurate results for classification and prediction
- Diagnose the disease in an early stage
- Affordable treatment for the patient
- Reduce doctor’s work

3.2 Flow Chart Model

The techniques used in this paper are represented by the graph bellow

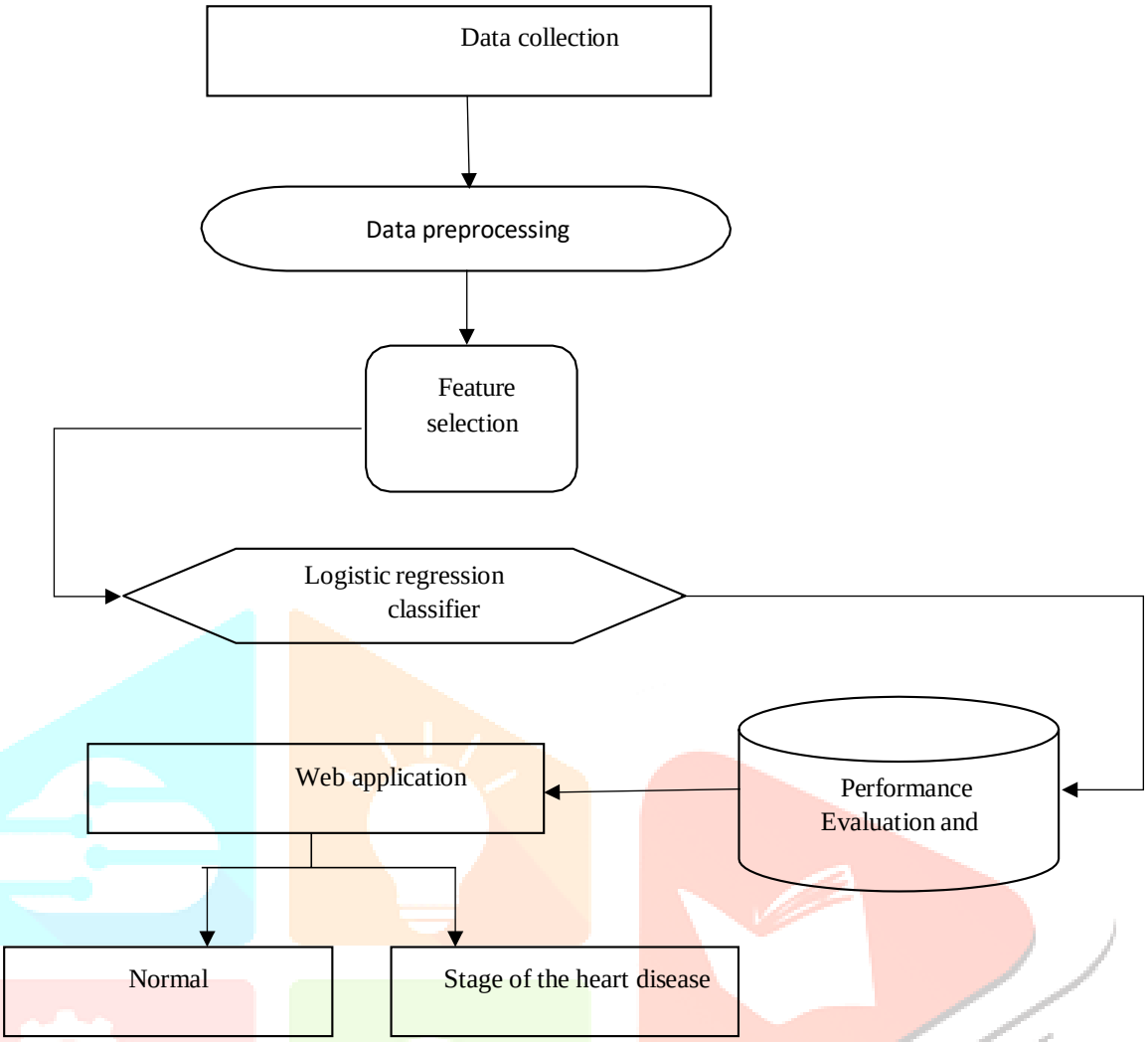

Fig 1 proposed model

Data Collection

This is the main step or the first step for all machine learning techniques. Data collection is the process of collecting or gathering all data from a specific or different source. The data used in this project is collected from the Kaggle repository. The dataset contains 604 rows and 14 columns. The 14 columns represent attributes listed in the bellow table 1

Table 1 Attribute Used

S.NO	Attributes	Description	Values
1	Age	The current age of the patient	Continuous value
2	Sex	Gender of the patient	0 = male 1 = female
3	Cp	Chest pain type	1 = typical Angina 2 = Atypical Angina 3 = Non-anginal pain 4 = Asymptomatic
4	Chol	Serum cholesterol	Continuous value
5	Threstbps	Resting blood pressure	Continuous values
6	Fbs	Fasting blood sugar	Value 0: <120mg/dl Value 1: >120 mg/dl
7	Restecg	Resting ECG Result	0= normal 1 = having ST-T wave abnormality

8	**Thalach**	**Max Heart Rate Achieved**	**Range [71, 202]**
9	**Exang**	**Exercise -Included Angina**	**0 = No** **1 = Yes**
10	**Oldpeak**	**ST depression included by exercise relative to rest**	**Continuous value**
11	**Slope**	**Slope of the peak exercise ST segment**	**Range in [1, 3] value**
12	**Ca**	**Number of Major Vessels**	**Range in [0, 3] value**
13	**Thal**	**Defect type**	**Value 7 = Reversible Defect** **Value 6 = Fixed Defect** **Value 7 = Normal**
14	**Target**	**Heart disease prediction**	**0 = normal** **1 = stage 1** **2 = stage 2** **3 = stage 3** **4 = stage 4**

Data Pre-processing

The data is imported into the python environment as a CSV file format. Independent variables such as (name, gender, age, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, exercise- induce angina, resting ECG result, max heart rate achieved, st depression, slop of the peak exercise st, number of major vessels) and dependent variables (target values) are extracted and stored as x and y respectively.

Feature Selection

It is the process of obtaining a subset from the original dataset without losing the features of the data set. In this step, irrelevant data, and noise data are removed. Removing irrelevant data provide a huge impact in the process such as improving the accuracy, reducing time, and easy understanding of the model. The data is divided into a training set and a testing set. The test set is used for scaling in order to get an accurate result for prediction. The training set is used to train the dataset.

Logistic Regression Classifier

Logistic regression is a supervised learning technique. It is one of the best machine learning algorithms. Logistic regression is getting very popular and used for classification and prediction due to its high accuracy. There are two types of regression models: binary logistic model and multinomial logistic regression model. In the Binary logistic regression model, the target variable can have either 1 or 0. in the other hand multinomial logistic regression model which is the model used in this project the dependent variable can have 3 or more possibilities.

The algorithm predicts dependent variables based on the independent variables. in this project, the logistic regression model predicts whether the patient has heart disease or not with the specific stage of the disease(target) based on the symptoms, personal details, and medical test (independents variable/attributes) of the user.

Fitting logistic regression to the training set:

- Import Logistic Regression class from the SK learn
- Create a classifier object
- Fit the model to the logistic regression
- Predicting the test result using predict method
- Import confusion matrix class to test the accuracy of the result
- A model is built and deployed into the web application

Web Application and Heart Disease Prediction

The web application is created using the flask web framework. The patient or the doctor can log in to the system and input the health attributes of the patient and click the predict button. The input given is compared with the existing model for prediction. After the prediction is done, the result is displayed to the user.

IV. SCREEN SHORTS OF HEART DISEASE PREDICTION SYSTEM

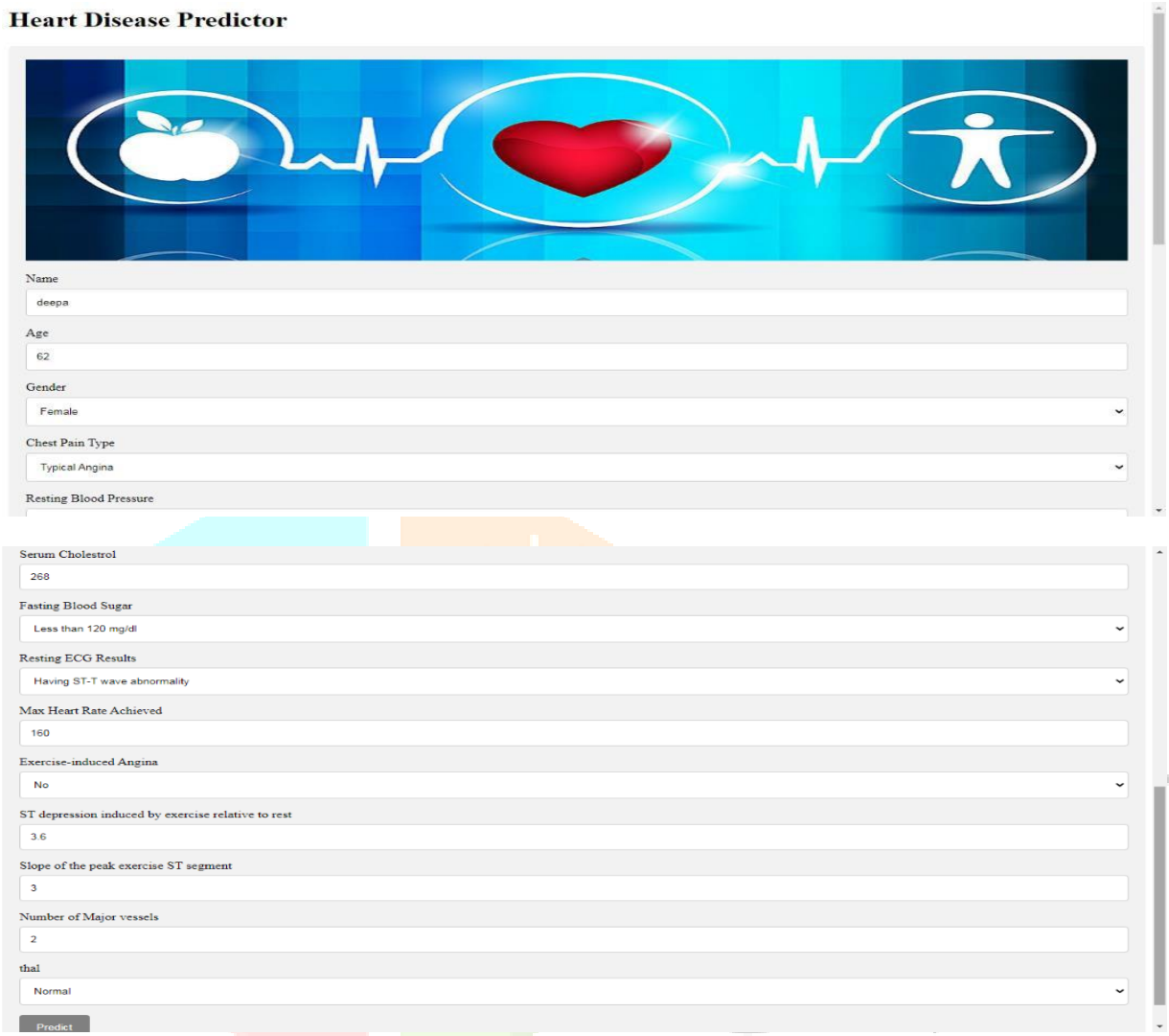

fig 2 patient details page

The above fig 2 represents the prediction page for the heart disease prediction system. This page represents the form of prediction that contains patient data. The input field is specified with a placeholder value for indication. When the doctor fills the form and clicks predict button, it will automatically redirect to the result page. The result page (fig 3) shown below will display the prediction result of whether the patient has heart disease or not with the specific stage of the disease and the patient report.

Prediction

Age	62
Gender	0
Chest Pain Type	1
Resting Blood Pressure	140
Serum Cholestrol	268
Fasting Blood Sugar	0
Resting ECG Results	1
Max Heart Rate Achieved	160
Exercise-induced Angina	0
ST depression induced by exercise relative to rest	3.6
Slope of the peak exercise ST segment	3
Number of Major vessels	2
Thal	3

NO! You might not have a heart Disease.

fig 3 Result Page

V. CONCLUSION AND FUTURE WORK

Heart disease prediction system is user-friendly, scalable, and reliable. It can be introduced in medical college as a tool for practice and research. The proposed model can help the doctor to make a better solution in an early stage with a low-cost treatment. The proposed machine learning model provides a sufficient and accurate result with a currency of 62%. The future enhancement of the heart disease prediction is to work to specify the particular CAD, CVD, and heart attacks. It can provide high-level accuracy results by including a decision tree classification algorithm in data mining and deep learning. This HDPS have potential in many fields such as industry, smart wear, hospital smartphone, and clinic. This model can be integrated with fitness mobile applications. Users can be integrated with hospital clinics using the HDPS model to predict heart disease. It can provide suggestions for precautions against heart disease when it's integrated with smart wear to detect heart disease. As early mentioned, this Project can deploy in a mobile application so that users can identify the heart disease by themselves. HDPS can integrate with hospital and police emergency systems for saving humans living in an emergency situation.

VI. REFERENCE

[1] Harshit Jindal et al (2021). Machine learning algorithms for heart disease prediction.

[2] Purushottam, Kanak Saxena, Richa Sharma. Efficient Heart Disease Prediction System, *Procedia Computer Science*, Volume 85, 2016, Pages 962-969, ISSN 1877-0509,

[3] Sunuwar, Ashish Adhikari et al. Heart Disease Prediction System. *International Journal of Software & Hardware Research in Engineering* 2021. 10.26821/IJSHRE.9.1.2021.9128

[4] Soni Jyoti. "Predictive datamining of medical diagnosis"

[5] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," 2008 IEEE/ACS International Conference on Computer Systems and Applications, 2008, pp. 108-115, DOI: 10.1109/AICCSA.2008.4493524.

[6] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December).

[7] [Online]. Available: https://www.kaggle.com/ronitf/heart-disease-uci#heart.csv...

[8] Bharti S 2015 International Conference on Computing, Communication and Automation (ICCA2015) (Noida) Analytical study of heart disease prediction comparing with different algorithms 78-82

Google Scholar

IJCRT