

Heart Disease Prediction using Logistic Regression Algorithm

Firoz Khan¹, Imran Khan², Nasreen Taj³, Raghavendra A⁴ and Chethan D R⁵

¹⁻³Assistant Professor, Department of Information science and Engineering, GM Institute of Technology, PB Road Davangere, Karnataka , India

Email: firozk@gmit.ac.in, imrank@gmit.ac.in, nasreent@gmit.ac.in

⁴⁻⁵Department of Information science and Engineering, GM Institute of Technology, PB Road Davangere, Karnataka , India

Email: nasreent@gmit.ac.in, raghavendraa623@gmail.com

Abstract—Cardiovascular disease causing deaths is a big problem in this new century. Approximately one person dies each minute from heart disease. Data is produced and needs to be processed daily, owing to the rapid growth of IT. By using different algorithm combinations the information obtained is converted by data analysis into knowledge. Health professionals who are working in the field of cardiac disease have their own limitations; they can not predict the risk of high precision cardiopathy. This paper aims to enhance cardiopathy predictive performance with the Machine Learning Logistic Regression Model taking into account the health care dataset that classifies patients whether they have heart disease or not using the recorded data.

Index Terms— Data Analysis, Coronary Heart Diseases, Logistic Regression Algorithm, sigmoid function.

I. INTRODUCTION

Over the last few years, the load of cardiovascular diseases has steadily increased across the world. Although these diseases were the most significant cause of death, they were known as the most manageable and preventable disease. Heart stroke induces primarily blockage in the arteries. It happens when heart doesn't pump the blood efficiently across the body.

Having hypertension is likewise one of the primary driver of getting a coronary illness. An overview says that, in 2011 to 2014, the ordinariness of hypertension on the planet was about 35%, which is additionally a reason for coronary illness. Additionally, there are a lot more explanations behind getting a coronary illness, for example, corpulence, not taking in legitimate nourishment, expanded cholesterol and absence of physical action. In this way, avoidance is important. For avoidance, familiarity with heart sicknesses is significant. Approximately 47 percent of individuals drop the bucket outside the medical centre and it shows they are not following up on signs of early warning.

A person's life expectancy is reduced these days as a result of heart infections. In 2013, the World Health Organization (WHO) established goals to combat non-transferable diseases (NCDs), along these lines, Of which 25% of the relative decrease is due to cardiovascular disease and it is assured that by 2025 at any point half of cardiovascular disease patients will seek substantial medicines and clinical guidance [1]. About 17.9 million

people kicked the bucket in 2016 because of cardiovascular diseases, which constitutes 31 per cent of the world's passages.

Identification of heart disease is an important examination. It's hard to guess whether a person has a coronary disease, or not. There are available tools that can predict heart disease, but they may be costly or unable to calculate the risk of heart disease in humans. According to a survey by the World Health Organization (WHO), clinical experts can only predict 67% of coronary disease, so there is tremendous scope for research in this field [2]. Access to major specialists and medical centres in country territories is extremely low if an Indian epidemic should occur. A study from the WHO in 2016 reports that only 58% of specialists have a professional degree in metropolitan areas and 19% in country zones.

In the U.S., someone has a clockwork-like respiratory failure, that is, more than one person passes on in the U.S. due to cardiovascular failure. Besides this, Turkmenistan has the highest speed of going through till 2012, with 712 passing for every 100,000 people. Though, Kazakhstan have the second most elevated pace of passing because of heart ailments.

In this arrangement India holds 56th position[3]. Additionally , the study reveals that 1.3 million cardiovascular moves, 0.9 million (68.4%) were caused by heart disease, and 0.4 million (28.0%) by stroke at 30-69 years of age. In clinical science, heart disease is a vital problem, and machine learning can be a good decision in humans to predict any cardiovascular disease. Usage of Neural Network, Decision Tree, KNN, etc., we can predict heart disease. Later in this paper we can see how Logistic Regression is used to figure out the exactness of coronary disease.

II. EXISTED SYSTEM

The before all current framework chips away at sets of both Deep learning and information mining. The current framework modules creates extensive report by actualizing the solid forecast calculation The primary points of the current framework to look at and check the before quiet whose having illness yields and new patient infection and decide future prospects of the coronary illness to a specific patient By Implementing the previously mentioned model we will get the objective of building up a framework with expanded pace of exactness of assessing the new patient getting cardiovascular failure rate. The model which is proposed for Heart Attack Prediction System is created for utilizing Deep learning calculations and approach. Be that as it may, by utilizing all the current frameworks the precision is extremely less.

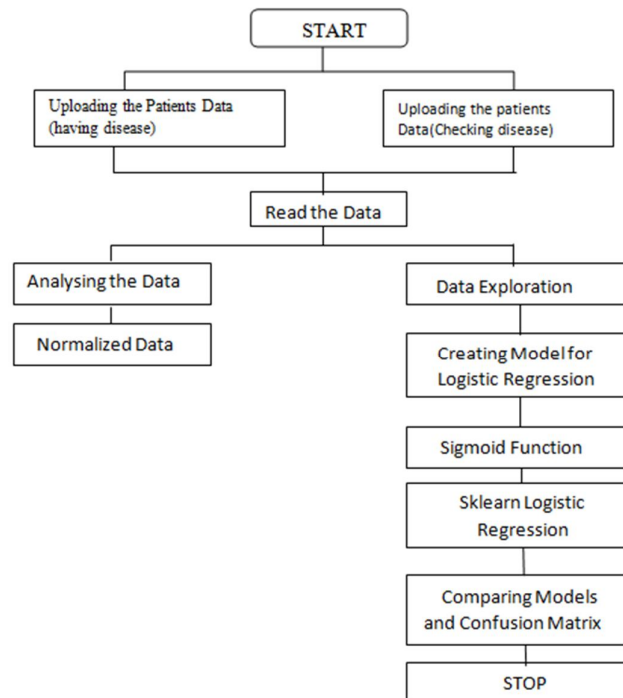


Figure 1 Flowchart of Proposed system

III. PROPOSED SYSTEM

This proposed framework has information which characterized if patients have coronary illness or not as indicated by highlights in it. This proposed framework can attempt to utilize this information to make a model which attempts anticipate (understanding information and information Exploration) if a patient has this sickness or not. In this proposed framework, utilize calculated relapse (grouping) calculation. By utilizing sklearn library to figure score. Actualizes Naive Bayes calculation to getting precision result. Using Matching Models and Uncertainty Matrix to help eventually analyse the outcomes. From the information we have it can be categorised into different ordered details based on the patient heart 's highlights. From the information's accessibility we need to build a model that predicts patient disease using measured relapse calculation. First we have to import the datasets. Peruse the datasets; the details should include different factors such as age, sexual orientation, sex, cp (chest pain), slope and goal. The details should be investigated with the aim of confirmation of the data. [4] Make an impermanent variable and furthermore fabricate a model for calculated relapse. Here, we utilize sigmoid capacity which helps in the graphical portrayal of the grouped information. By utilizing calculated relapse, gullible byes the precision rate increments.

IV. METHODOLOGY AND APPROACH

The World Health Organization's presentation has measured 12 million happenings around the world; regularly related to heart disease. Cardiovascular disorders in the United States and other produced nations are estimated to account for a large portion of the passage. In high-risk patients, early awareness of cardiovascular disorders may help decide on life-changing decisions and thereby lessen the entanglements. This study expects the most applicable / chance variables of coronary disease to be defined just as the general threat using measured Data Planning relapse is expected. Strategic Regression is a sort of relapse examination in insights utilized for forecast of result of an unmitigated ward variable from a lot of indicator or free factors. In strategic relapse the needy variable is consistently twofold. Strategic relapse is for the most part used to for expectation and furthermore figuring the likelihood of success [5].

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sys
count	4238.000000	4238.000000	4133.000000	4238.000000	4209.000000	4185.000000	4238.000000	4238.000000	4238.000000	4188.000000	4238.000
mean	0.429212	49.584946	1.978950	0.494101	9.003089	0.029630	0.005899	0.310524	0.025720	236.721585	132.352
std	0.495022	8.572160	1.019791	0.500024	11.920094	0.169584	0.076587	0.462763	0.158316	44.590334	22.038
min	0.000000	32.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	107.000000	83.500
25%	0.000000	42.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	206.000000	117.000
50%	0.000000	49.000000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	234.000000	128.000
75%	1.000000	56.000000	3.000000	1.000000	20.000000	0.000000	0.000000	1.000000	0.000000	263.000000	144.000
max	1.000000	70.000000	4.000000	1.000000	70.000000	1.000000	1.000000	1.000000	1.000000	696.000000	295.000

Figure 2 Distribution of a Dataset

The dataset as shown in Fig.2 is from ongoing cardiovascular study of the people of Framingham town of Massachusetts. The assessment goal is to determine if the patient is at risk of possible coronary heart disease (CHD) for a duration of 10 years.

	male	age	cigsPerDay	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	TenYearCHD
0	1	39	0.0	0	0	0	195.0	105.0	70.0	26.97	80.0	77.0	0
1	0	46	0.0	0	0	0	250.0	121.0	81.0	28.73	95.0	75.0	0
2	1	48	20.0	0	0	0	245.0	127.5	80.0	25.34	75.0	70.0	0
3	0	61	30.0	0	1	0	225.0	150.0	95.0	28.58	65.0	103.0	1
4	0	46	23.0	0	0	0	285.0	130.0	84.0	23.10	85.0	85.0	0
5	0	43	0.0	0	1	0	228.0	180.0	110.0	30.30	77.0	99.0	0
6	0	63	0.0	0	0	0	205.0	138.0	71.0	33.11	80.0	85.0	1
7	0	45	20.0	0	0	0	313.0	100.0	71.0	21.68	79.0	79.0	0
8	1	52	0.0	0	1	0	260.0	141.5	89.0	26.36	78.0	79.0	0
9	1	43	30.0	0	1	0	225.0	162.0	107.0	23.61	93.0	88.0	0

Figure 3 Dataset after Wrangling

	precision	recall	f1-score	support
0	0.99	0.84	0.91	716
1	0.09	0.69	0.16	16
accuracy			0.84	732
macro avg	0.54	0.77	0.54	732
weighted avg	0.97	0.84	0.90	732

Figure 4 Accuracy Result

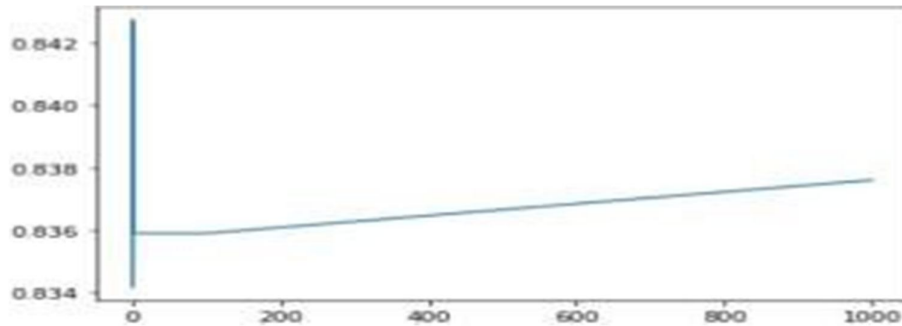


Figure 5 Accuracy Vs Alpha Result

V. RESULTS

From the above measurements plainly the model is profoundly explicit than delicate Men are more vulnerable to coronary illness than women. Increase in age, amount of cigarettes smoked daily, and in addition, systolic blood pressure suggests elevated risk for heart disease. Absolute cholesterol shows no substantial increase in the odds of developing CHD. This may be due to the presence of healthy cholesterol (HDL) in the absolute perusing cholesterol. The model predicted a precision of 0.84 as can be seen in Fig.3. The model is more articulate than touchy. The model could be developed by and large with more data and using more Machine Learning methods.

VI. CONCLUSION

The Heart disease calculation will go beyond the current situation to reach the most severe stage. Coronary disease is uncertain, with bunches of individuals hit the bucket with this infection every single year. The risks of having coronary disease based on danger factors have recently appeared to be difficult to physically determine. By using this approach, one of the major limitations of this study is that its main core is primarily focused on the use of arrangement methods and estimates for predictions of coronary disease, Considering various information cleaning and mining techniques to plan and generate a data set suitable for information mining, so that we can use this Machine Learning in calculated relapse calculations by predicting whether coronary disease is tolerant or not. Any non-clinical worker may use this tool to predict coronary disease and reduce the specialists' time-faced existence. It is still an open field that holds back for actualization in predication of coronary disease and improves precision.

REFERENCES

- [1] The Lancet Global Health. The changing patterns of cardiovascular diseases and their risk factors in the states of India: The global burden of disease study 1990-2016.
- [2] Himanshu Sharma, M A Rizvi. Prediction of heart disease using machine learning algorithms: A survey.
- [3] Sana Bharti, 2015. Analytical study of heart disease prediction compared with different algorithms; International conference on computing, communication, and automation (ICCA2015).
- [4] Gritsenko, Elena. "Health Care Analytics: Modelling Behavioural Risk Factors Associated With Disease." (2019).
- [5] Tarun Saxena, Rahul Yadav . "Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning" (2019)

BIOGRAPHICAL NOTES

Mr. Firoz khan Assistant Professor, GM Institute of Technology with a 10 years of experience in teaching and doing his research in the field of Medical Image Processing.

Mr. Imran khan Assistant Professor, GM Institute of Technology with a 10 years of experience in teaching and doing his research in the field of Medical Image Processing

Mrs. Nasreen Taj Assistant Professor, GM Institute of Technology with a 10 years of experience in teaching and doing his research in the field of Medical Image Processing

Mr.Raghavendra A is a Data Science enthusiast pursuing Bachelor of Engineering at GM Institute of Technology in Information Science and Engineering domain , Davangere, Karnataka, India .

Mr. Chethan D R is a Data Science enthusiast pursuing Bachelor of Engineering at GM Institute of Technology in Information Science and Engineering domain , Davangere, Karnataka, India.