Building a Heart Disease Prediction System Using Machine Learning Algorithm

Nikky George Philip
Department of Computer Applications
Amal Jyothi College of Engineering, koovapally
Kottayam, Kerala.
nikkygeorgephilip@mca.ajce.in

Mr. Jinson Devis
Asst. Professor in Computer Application
Amal Jyothi College of Engineering, koovapally
Kottayam, Kerala.
jinsondevis@amaljyothi.ac.in

***Abstract*— Heart disease can be considered as a serious issue that affect our hearts. Heart related diseases are increasing day by day that may lead to death also. Heart disease occurs when a waxy substance known as plaque develops in the arteries and blood vessels that lead to the heart. Thus the plaque blocks the important nutrients, oxygen from reaching our heart. Plaque is a waxy substance developed as a combination of cholesterol, fat molecules, and some minerals. It accumulates over time when the inner lining of an artery is damaged by high blood pressure, cigarette smoking, or elevated cholesterol or triglycerides. High blood pressure can be considered as one of the main reasons. If this high blood pressure is not controlled it may affect the heart and other organs in our body. In this paper we discuss how Machine Learning can be applied and implemented to build a prediction system for detecting heart disease in people.**

Keywords—heart disease prediction, machine learning algorithm, logistic regression.

I. INTRODUCTION

We know that in our body heart plays an important role in the circulatory system. Heart is responsible for the supply of blood to all other organs of our body. The blood that is pumped by the heart carries oxygen and nutrients to the body and carries metabolic waste such as carbon dioxide to the lungs. If the heart stops working suddenly all other organs of our body stop working altogether. There may be many situations that the heart can't function well. One of the main reasons for the improper function of the heart may be the block that forms inside the blood vessels. If such situations happen then the circulation of blood to other organs of the body through the blood vessels stops working, death occurs within minutes. Cardiovascular disease is defined as the state of blocked or narrowed blood vessels which leads to chest pain, heart attack, or stroke. The symptoms of heart disease in men and women may be different. The mechanisms vary looking on the illness. arteria coronaria illness, stroke, and peripheral artery illness involve induration of the arteries. this could be caused by high-pressure levels, daily cigarette smoking, lack of workout, heavy alcohol usage, and lack of sleeping, in others. This may be considered as a danger situation that when the heart stops working the circulation of blood to all organs of the body will be stopped. If we didn't take proper precautions and medicines heart disease can lead to death.

II. LITERATURE REVIEW

Senthil Kumar Mohan et al,[1] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which the system is the goal is to find basic incorporates by applying Machine Learning achieving working on the precision in the assumption for cardiovascular illness. The assumption model is made with different mixes of features and a couple of known course of action procedures. We produce a further developed presentation level with an accuracy level of 88.7% through the forecast model for coronary illness with mixture irregular woods with a direct model (HRFLM) they moreover taught about Diverse information mining approaches and assumption procedures, Such as KNN, LR, SVM, NN, and Vote have been genuinely well known of late to recognize and foresee coronary illness.

Sonam Nikhar et al [2] have developed the paper named Prediction of Heart Disease Using Machine Learning Algorithms by This investigation intends to give a point-by-point depiction of Naïve Bayes and choice tree classifiers that are applied in our assessment particularly in the expectation of Heart Disease. Some examination has been directed to contemplate the execution of insightful information mining methodology on the identical dataset, and the outcome reveals that Decision Tree beats over Bayesian grouping framework.

Abhay Kishore et al,[3] created Heart Attack Prediction Utilizing Deep Learning in which This paper proposes a heart assault forecast framework utilizing Deep learning methodology, unequivocally Recurrent Neural System to anticipate the plausible possibilities of heart related contaminations of the patient. Intermittent Neural Network is a very earth shattering portrayal estimation that uses Deep

DOI: 10.5281/zenodo.6181407
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering

Learning approach in Artificial Neural Network. The paperdiscusses exhaustively the huge modules of the system closeby the connected speculation. The proposed model profoundlearning and information mining to give the exact results least
goofs. This paper gives a direction and perspective for the progression of one more sort of coronary episode forecast stage. Forecast stage.

III. MACHINE LEARNING ALGORITHM

Machine learning is the way of analyzing the given information that considers automating the build of analytical model. It's associates with artificial intelligence supported the thought that systems will have the ability to learn from data provided, determine patterns and create selections with terribly minimum amount of human interaction. With the introduction of new computing technologies, the current day machine learning has emerged as an efficient field however it had been in its origin. Machine learning follows the pattern recognition method so that it is not considered to be programmed for specific tasks. Machine learning learn the given set of data and will be able to predict the outcome of the next given data. Previous computations are learned to obtain a meaningful result. Machine learning may be considered to be the study of computing technology with the help of computer science that specialize in the information and algorithms that follows the way in which humans learns something with high accuracy. Th medical field is considered to be one of the main streams where the application of machine learning takes place more efficiently to predict the results of the data records. With the efficient advancements in machine learning we can easily detect and predict heart disease without the intervention of human. In Machine Learning Classification is the style of dividing the information learned by the algorithmic program into 2 categories in our dataset. The categories are often one or zero and 'Yes' or 'NO'.

IV. PROPOSED METHODOLOGY

The main purpose of this paper is to build a Heart Disease prediction system using the Logistic Regression model with the health record dataset which classifies the patients whether they are having heart diseases or not according to the information given in the record. The model should be able to correctly classify the provided dataset into Heathy and Heart Diseased groups. The dataset provided is mainly burst into a set of training data and a set of testing data. A dataset with a limited number of data can't predict the correct accuracy rate so we are training the model using more data. The most vital step in this study is to create a predicting system. The value of a Heart Diseased entry will be 1 else 0 if the classification model can predict correctly. Here we use the logistic regression model for along with the training data. The flow of our study is as follows:

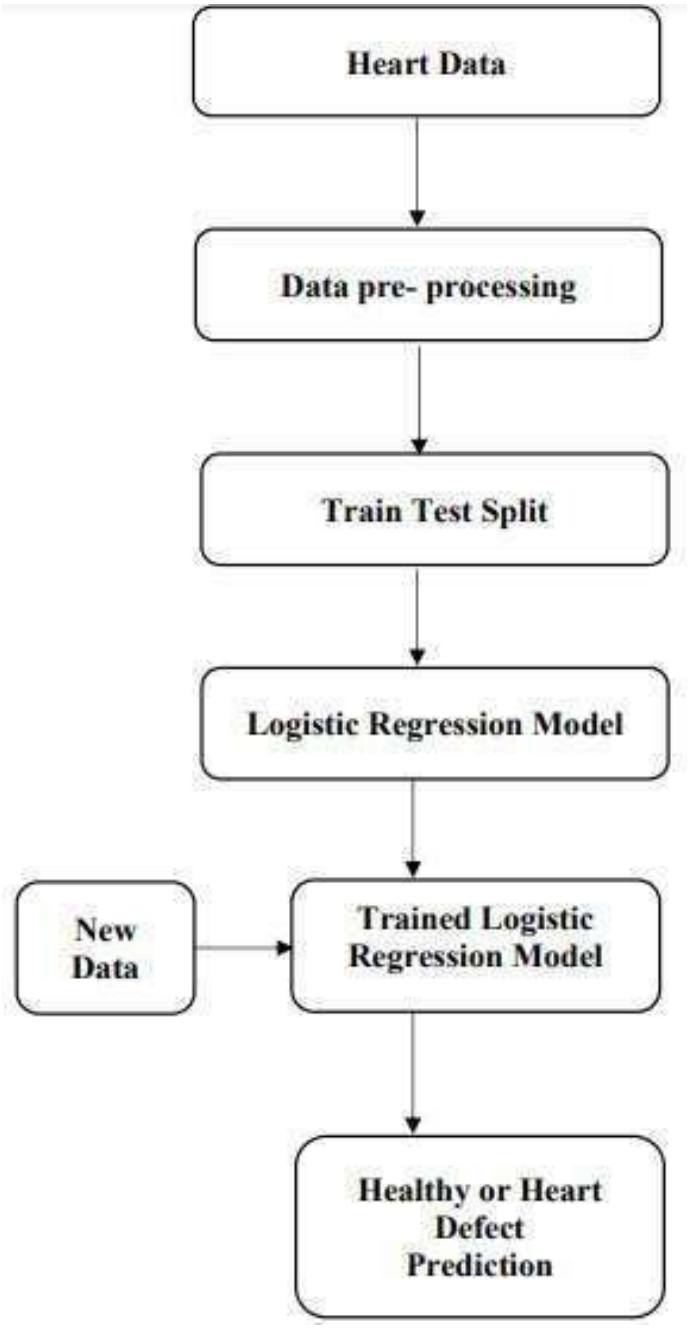

Figure 1 : Heart Disease prediction system data flow

The steps we follow are:

1. Importing the dependencies.
2. Data Collection and Processing datasets.
3. Splitting features and target.
4. Separating the data into two groups: training and test.
5. Model Training (Logistic Regression Model).
6. Model Evaluation.

DOI: 10.5281/zenodo.6181407
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering

7. Building a Predictive System.
8. Inputting new data to the trained model.

The Heart Dataset from the UCI Repository is the dataset we used in this study. The following are the attributes in this dataset:

Table 1: Dataset Description

SL.No	Attributes
1	age
2	sex
3	chest pain type
4	resting blood pressure
5	serum cholestoral in mg/dl
6	fasting blood sugar > 120 mg/dl
7	resting electrocardiographic results (values 0,1,2)
8	maximum heart rate achieved
9	exercise induced angina
10	oldpeak = ST depression induced by exercise relative to rest
11	the slope of the peak exercise ST segment
12	number of major vessels (0-3) colored by flourosopy
13	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

The next attribute in the dataset is the class variable having 1 or 0 as the value. This variable represents the case whether a person is healthy or he is having heart disease. In this study, we are using around 300 data records that are imbalanced and out of that 165 are labelled as 1 which represents they are having heart disease and 138 are labelled as 0 which represents they are healthy.

A. Data pre-processing: This is considered to be the first step of the learning process. In this step we remove the missing values and all impurities from our dataset and make the data good for processing. The purpose of data pre-processing is to improve the accuracy of the data that will be processed. The following methods can be used to accomplish data pre-processing:

1. **Removing missing values:** A dataset can contain missing values suck as 0 for some entries. We have to remove those values from the dataset and make the dataset for efficient processing.

2. **Splitting the dataset:** Once the impurities are removed from the dataset we have to split the given dataset into a category of Training and Testing sets. The model is trained using the training set.

Logistic Regression Model: For classification, the Logistic Regression model is considered to be the best option. The most common machine learning algorithm is logistic regression. It has got a high level of accuracy and is usually very definitive. 0s and 1s are the results that are usually noted with logistic regression. It is normally used when data needs to be categorized. To find the probability of the valid class and invalid class we use the sigmoid function.

Sigmoid function P = 1/1+e-(a+bx)

Here P = probability, a and b are model parameters.

V. BUILD THE MODEL

Model building is the key stage in the prediction of heart disease. In this stage we implement the Logistic Regression model we have discussed above:

Steps:

1. Importing the libraries.
2. Importing the heart dataset.
3. Data pre-processing is used to remove missing values from the dataset.
4. Separating the data into two groups: training and test in a 80:20 ratio.
5. Selecting the Logistic regression Algorithm.
6. Training the model using the algorithm with training data.
7. Evaluating the accuracy of the trained model using the training data and testing data.
8. Building a predictive system.
9. Using data to evaluate the prediction system's performance.

VI. RESULTS

We have evaluated the performance of the Machine Learning Algorithm for predicting Heart Disease. The classifier model has given following accuracies on predicting the training data and test data.

Accuracy Score

```
# accuracy on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```

```
print('Accuracy on Training data : ', training_data_accuracy)
```

```
Accuracy on Training data :  0.8512396694214877
```

```
# accuracy on test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
print('Accuracy on Test data : ', test_data_accuracy)
```

```
Accuracy on Test data :  0.819672131147541
```

DOI: 10.5281/zenodo.6181407
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering

Figure 2: Accuracy results of the Logistic model.

Result of the predictive system is as follows:

```
input_data = (62,0,0,140,268,0,0,160,0,3.6,0,2,2)

# change the input data to a numpy array
input_data_as_numpy_array= np.asarray(input_data)

# reshape the numpy array as we are predicting for only on instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if (prediction[0]== 0):
  print('The Person does not have a Heart Disease')
else:
  print('The Person has Heart Disease')
```

```
[0]
The Person does not have a Heart Disease
```

Figure 3: Result of the Prediction system.

VII. CONCLUSION

With this research, we have applied a machine learning algorithm called logistic regression in order to predict heart disease from the patients' data set. With the help of this study, we were able to build a system for predicting heart disease.

VIII. REFERENCES

[1] [1] Senthilkumar Mohan, Chandrasegar Thirumalai, GautamSrivastava ―Effective Heart Disease Prediction Using HybridMachine Learning Techniques‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, ―On the Compatibility of Adaptive Controllers,‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)

[2] [2] Sonam Nikhar, A.M. Karandikar" Prediction of Heart Disease Using Machine Learning Algorithms" International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication,[Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.Devansh Shah, Samir Patel & Santosh Kumar Bharti , Heart Disease Prediction using Machine Learning Techniques , SN Computer Science volume.

[3] Abhay Kishore1, Ajay Kumar2, Karan Singh3, Maninder Punia4, Yogita Hambir5," Heart Attack Prediction Using Deep Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.

[4] Santhana Krishnan J.; Geetha S., Prediction of Heart Disease Using Machine Learning Algorithms , 10.1109/ICIICT1.2019.8741465

[5] Harshit Jindal , Sarthak Agrawal , Rishabh Khera , Rachna Jain Heart disease prediction using machine learning algorithms , January 2021, IOP Conference Series Materials Science and Engineering 1022(1):012072 , DOI:10.1088/1757-899X/1022/1/012072.

[6] Mursal Furqan , Hiba Rajput , Sanam Narejo , Adnan Ashraf , Heart Disease Prediction using Machine Learning Algorithms

[7] Harshit Jindal , Sarthak Agrawal , Rishabh Khera , Rachna Jain Heart disease prediction using machine learning algorithms , January 2021, IOP Conference Series Materials Science and Engineering.

DOI: 10.5281/zenodo.6181407
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering