HEART DISEASE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING

Dinesh Kalla and Arvind Chandrasekaran

Department of Computer Science, Colorado Technical University, Colorado, USA

ABSTRACT

Heart disease is most common disease reported currently in the United States among both the genders and according to official statistics about fifty percent of the American population is suffering from some form of cardiovascular disease. This paper performs chi square tests and linear regression analysis to predict heart disease based on the symptoms like chest pain and dizziness. This paper will help healthcare sectors to provide better assistance for patients suffering from heart disease by predicting it in beginning stage of disease. Chi square test is conducted to identify whether there is a relation between chest pain and heart disease cases in the United States by analyzing heart disease dataset from IEEE Data Port. The test results and analysis show that males in the United States are most likely to develop heart disease with the symptoms like chest pain, dizziness, shortness of breath, fatigue, and nausea. This test also shows that there is a week corelation of 0.5 is identified which shows people with all ages including teens can face heart diseases and its prevalence increase with age. Also, the tests indicate that 90 percent of the participant who are facing severe chest pain is suffering from heart disease where majority of the successful heart disease identified is in males and only 10 percent participants are identified as healthy. The evaluated p-values are much greater than the statistical threshold of 0.05 which concludes factors like sex, Exercise angina, Cholesterol, old peak, ST_Slope, obesity, and blood sugar play significant role in onset of cardiovascular disease. We have tested the dataset with prediction model built on logistic regression and observed an accuracy of 85.12 percent.

KEYWORDS

Chi-Square Test, R; Data Mining; Big Data; Linear Regression Analysis; Heart Disease; Risk Factor; Machine Learning; Cardiovascular Disease; Python; Logistic Regression; sklearn; Pandas, Numpy, NLTK.

1. INTRODUCTION

Cardiovascular disease describes various conditions that can affect the human heart. Heart disease is among the deadliest and the most complex human disease across the globe [1]. Accordig to the reports from World Health Organization (WHO), cardiovascular disease kills 17.9 million people per year globally. [9] claims that in heart disease, the heart pumps insufficient amount of blood to other body organs which affects their functionalities. According to [2], some of the activities that increase the likelihood of developing heart disease are obesity, high levels of cholesterol, high blood pressure among others. In addition, age, genetic and past events also influence the likelihood of developing heart diseases [5]. As described by American Heart Association, individuals suffering from heart disease show various signs and symptoms. These persons experience challenges in their sleep, irregular heartbeat (heart rate decrease or increase), rapid weight loss and swollen legs. However, these signs and symptoms are common for different diseases especially in aged persons. Therefore, it is difficult to get the actual diagnosis which may lead to increased mortality in the near future.

DOI:10.5121/ijdkp.2023.13301

Correct diagnosis of heart disease is critical towards reducing the mortality. Prediction helps Physicians often use angiography approach to diagnose heart disease. However, this diagnostic approach is time consuming and cost ineffective especially in developing countries where healthcare providers, diagnostic technologies among other resources are limited [3]. In recent years, health industry is incorporating modern technology to offer better services to patients. With the modern technological advances in health industry, patient’s data can easily be accessed through several available open sources. Using this data, researches can be carried out so that different modern technologies can be used to correctly diagnose patients and detect heart disease before the condition worsen [5]. Artificial intelligence and machine learning are playing critical in the prediction and detection of heart diseases. Different models of deep learning and machine learning can be used to diagnose cardiovascular disease and predict the outcomes [9]. Researchers use different machine learning to conduct a comprehensive genomic data analysis within short time and with high accuracy degree.

Traditional diagnostic methods for heart disease typically involve invasive techniques that rely on a comprehensive evaluation of the patient's medical history, physical examination, and a thorough analysis of the patient's symptoms by medical professionals [3]. Despite the significant advancements in medical science and technology, these traditional methods still have inherent limitations, including inaccuracies and delays in diagnosis results, which can be attributed to human error. Furthermore, the use of these traditional diagnostic methods often requires a significant amount of financial resources, as well as advanced computational and technical expertise, and can also be time-consuming, leading to additional stress and anxiety for patients.

This paper will analyse a dataset containing information about five different heart diseases. The data set is representative of a single large data set on cardiovascular disease thanks to the inclusion of twelve standard features. Researchers can use methods like machine learning on the dataset to learn more about the trend, identify the most at-risk populations, and discover other insights. This study will make use of visual representations to learn about the dataset. This will help the health ministry better provide care for patients suffering from heart disease by predicting the earliest stages of the disease. The dataset we are using to build heart prediction model contains 303 rows and 14 columns will all health-related data. We will use same dataset to check the prediction algorithm accuracy which is developed using logical regression.

2. RELATED WORK

A lot of effort has been put into developing disease prediction systems in hospitals, mostly utilizing data mining and machine learning. Heart disease prediction allows healthcare providers make informed decisions about the health of a patient. Using machine learning helps to understand and reduce the symptoms of cardiovascular diseases. Heart Disease can be predicted utilizing the Multiple Regression Model, demonstrating the validity of Multiple Linear Regression [13]. The work is done on the data set of 3000 instances with 13 different attributes. 70% of the data is used for training purposes, while the remaining 30% is used for validation. Regression's classification accuracy is higher than competing algorithms, as demonstrated by the results.

Mangione's research team has developed an innovative heart disease prediction model that leverages advanced algorithms, including KStar, Multilayer perceptions, SMO, Bayes Net, and J48 [7]. While these algorithms show promise in accurately predicting heart disease, they have not yet achieved optimal levels of accuracy. To improve the accuracy of the model, the research team has implemented fold cross-validation techniques, which have proven to be more effective than traditional methods of validation. This has resulted in an increase in the overall performance of the model in terms of accuracy, which is essential for better decision-making in disease diagnosis. The improved accuracy of the heart disease prediction model can significantly reduce the number of deaths caused by heart disease, thereby improving public health outcomes.

In recent years, data mining techniques have been widely used in healthcare to predict and diagnose chronic diseases based on previous health records. One such study by researchers proposed using Decision tree, Support Vector Machines (SVM), Naive Bayes, and Artificial Neural Networks (ANN) to predict chronic diseases by mining data in previous health records [11].To determine the classifier with the highest accuracy rate, the researchers conducted a comparison study. The results of the experiment showed that SVM had the highest accuracy rate, making it an effective method for predicting chronic diseases. Additionally, the study found that Naive Bayes was the best method for diagnosing diabetes, suggesting that different algorithms may be more effective for different types of chronic diseases.By leveraging data mining techniques, healthcare professionals can gain valuable insights into a patient's health history, enabling them to make more informed decisions about diagnosis, treatment, and prevention.

Different algorithms, for example, Naive Bayes, Classification Tree, ANN, SVM, and Logistic Regression can be used in predicting cardiovascular diseases [8]. Compared to other algorithms, Logistic Regression has the highest level of precision. NLP can also be used to develop medical chatbots to help patients suffering from breast cancer or heart disease [6].

When working with machine learning, dealing with high dimensionality of data is a common challenge. With datasets that contain huge amounts of data, it can be difficult to even visualize the data in three dimensions, which is referred to as the curse of dimensionality. The processing of such large datasets can require a huge amount of memory and can lead to issues such as overfitting. However, one approach to addressing this issue is to use weighting features, which can decrease the redundancy in the dataset and reduce the processing time required for execution. To further tackle the dimensionality of the dataset, there are various techniques for feature engineering and feature selection that can be utilized to remove data that may not be as important in the overall dataset.

The use of data mining techniques has emerged as a valuable tool for predicting and diagnosing various diseases, including cardiovascular disease. [10]. Two of WEKA's many applications are automatic disease diagnosis and service quality assessment in healthcare facilities. SVM, ANN, Naive Bayes, Association rule, and Decision Tree were among the algorithms utilized in the paper. The research paper concludes that SVM is the most efficient and accurate algorithm for predicting cardiovascular disease. This is due to the ability of SVM to handle complex and high-dimensional data, making it well-suited for medical datasets. Additionally, SVM can efficiently identify the most important features and attributes for prediction, enabling healthcare professionals to make better decisions regarding treatment and care.

Data mining is a powerful tool that can be used to analyse large datasets and extract useful information. In the healthcare industry, prediction and the Analysis of Heart Diseases Incidence can be achieved by utilizing Data Mining Method [5]. The primary goal is to automate the early diagnosis of heart disease by predicting when it will occur. In healthcare organizations, the proposed methodology is also crucial for dealing with experts who have lost their expertise. The presence or absence of heart disease is determined using different medical characteristics, for instance, blood sugar, age, and heart rate. By utilizing data mining algorithms, healthcare organizations can create a repository of knowledge that can be used to diagnose heart disease and other medical conditions.

Non-linear classification algorithms have been gaining popularity in predicting cardiovascular disease due to their ability to handle complex relationships among various predictors. Non -linear classification algorithms can be utilized and adopted for predicting cardiovascular disease [5]. In particular, big data tools such as Hadoop Distributed File System (HDFS), MapReduce, and Support Vector Machines (SVM) can be utilized to achieve accurate predictions This study examined how various data mining strategies could foretell cardiac conditions. It recommends using HDFS to distribute large datasets across multiple nodes, each of which can run the SVM prediction algorithm in parallel. The computation time for SVM is reduced by using it in a parallel fashion rather than in a sequential fashion.

Data mining and machine learning algorithms together can be used for predicting cardiovascular diseases [12]. The research purpose is to reveal previously hidden relationships using data mining methods. Gomathi et al. proposed using data mining methods for multi-disease prediction. Data mining is now an essential tool for diagnosing many diseases because it allows for a decrease in the number of necessary evaluations. The primary focus of this paper was predictions of chronic diseases like diabetes and cancer.

ANN algorithm can be utilized in data mining for heart disease prediction in all stages [4]. The rising cost of diagnosing cardiovascular disease has prompted the search for a more cost-effective method of early detection. After collecting data on vital signs like heart rate, blood pressure, and cholesterol, the prediction model can be used to estimate how the patient will fare in the future. Verification of the system's accuracy has been implemented in JavaScript. Machine Learning and Artificial Intelligence need to be incorporated into application to improve the performance of the Application [3]. Natural Language processing library plays significant role in building machine learning models [4]. Surveys need to be conducted to build a dataset which can used further for designing logistic regression model [13].

Zhou's team proposed developing a prediction system to detect heart disease from patients' medical records [14]. The system was created with 13 input attribute risk features in mind. Different data mining algorithms, including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR), were used to develop the heart disease prediction system. The performance of each algorithm was evaluated based on different metrics, including accuracy, precision, recall, and F1-score. Data cleaning and integration followed the analysis of the dataset's information.

2.1. Design Model

To develop prediction model heart dataset from IEEE port is used and the dataset contains several health parameters. In the first stage dataset will beprocessedand split into train and test data. In the second stage the dataset is used to train the machine learning algorithm using the training data. Finally same initial dataset is used to test the machine learning algorithm model with the test data.

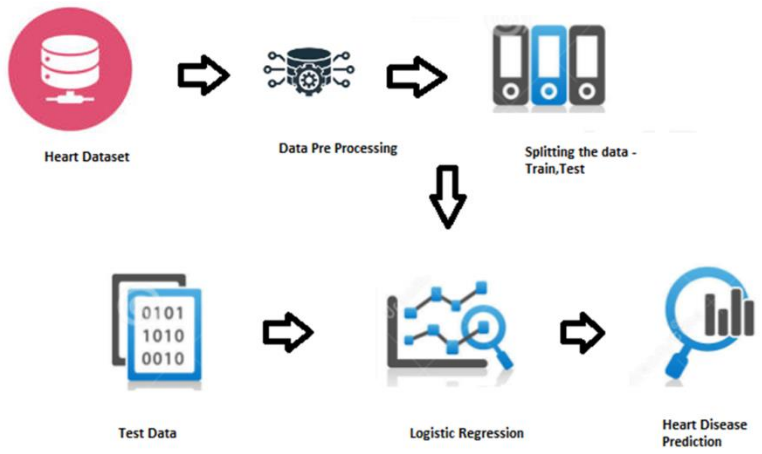

Figure 1. Logistic Regression Model using Machine Learning Algorithm

In this prediction model logistic regression model is used as it needs to do binary classification. After training the regression model it will pass the new dataset and will check whether the person has heart disease or not. After comparing the dataset result, it will measure the accuracy of the machine learning algorithm.

2.2. Data Collection

Chou's team compiled the data on cardiovascular disease for primary prevention of cardiovascular disease [3]. The primary goal of this data collection is to give scientists a large enough sample to work with to improve machine learning and data mining techniques, leading to better methods for diagnosing and treating heart disease. As can be seen in the table below, the dataset contains 1191 instances and 12 distinct attributes.

```
> attributes(heart_statlog_cleveland_hungary_final)
$names
 [1] "age"               "sex"              "chest pain type"
 [4] "resting bp s"      "cholesterol"      "fasting blood sugar"
 [7] "resting ecg"       "max heart rate"   "exercise angina"
[10] "oldpeak"           "ST slope"         "target"
```

The dataset is one of the large heart disease data sets because it contains 12 typical features. Researchers can use the dataset and methods like machine learning to learn more about the trend, identify the most at-risk populations, and so on.

Table 1. Description of Heart Disease Dataset Attribute

Serial Number	Attribute	Code Given	Unit	Data Type
1	Age	Age	In years	Numeric
2	Sex	Sex	1,0	Binary
3	Chest Pain Type	Chest Pain Type	1,2,3,4	Nominal
4	Resting Blood Pressure	Resting bp s	In mm Hg	Numeric
5	Serum Cholesterol	Cholesterol	In mg/dl	Numeric
6	Fasting blood sugar	Fasting blood sugar	1,0>120 mg/dl	Binary

7	Resting electrocardiogram results	Resting ecg	0,1,2	Nominal
8	Maximum heart rate achieved	Max heart rate	71-202	Numeric
9	Exercise induced angina	Exercise angina	0,1	Binary
10	Oldpeak=ST	Oldpeak	depression	Numeric
11	The slope of the peak exercise ST segment	ST slope	0,1,2	Nominal
12	class	target	0,1	Binary

Table 2. Nominal Attributes Description

Attributes	**Description**
Sex	Male =1, Female = 0
Chest Pain Type	--Value 1: typical Angina --Value 2: atypical Angina --Value 3:non- Angina pain --Value 4:asymptomatic
Fasting Blood Sugar	Fasting blood sugar>120mg/dl (True=1, False =0)
Resting electrocardiogram results	--Value 0: normal --Value 1: having ST-T wave abnormality (ST Elevation / depression >0.05 mV /T-wave Inversions) --Value 2: Definite ventricular left side hypertrophy by Estes criteria or showing probable
Exercise induced angina	Yes = 1, No = 0
The slope of the peak exercise ST segment	--Value 1: upsloping --Value 2: flat --Value 3:down sloping
Class	Heart Disease = 1, Normal = 0

2.3. Exploratory Data Analysis

EDA performs a critical analysis to comprehend better data patterns and aid in detecting anomalies using visual representations of data, such as a histogram.

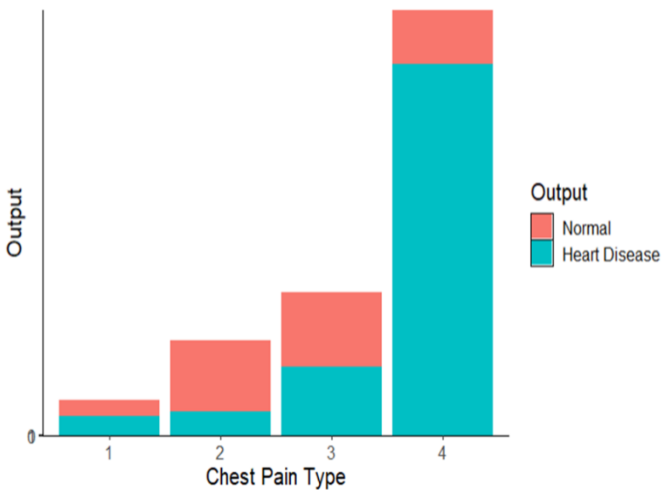

Figure 2. Chest Pain Types Histogram

Above figureshows data on chest pain has been graphically represented as a histogram. It is clear from Fig. 1 that there are two types of variables: categorical variables, such as the types of chest pains, and continuous variables, such as the severity of the pain. Ninety percent of the volunteers had heart disease, while 10 percent were healthy—more males than females reported experiencing symptoms of heart disease (chest pain) in this study.

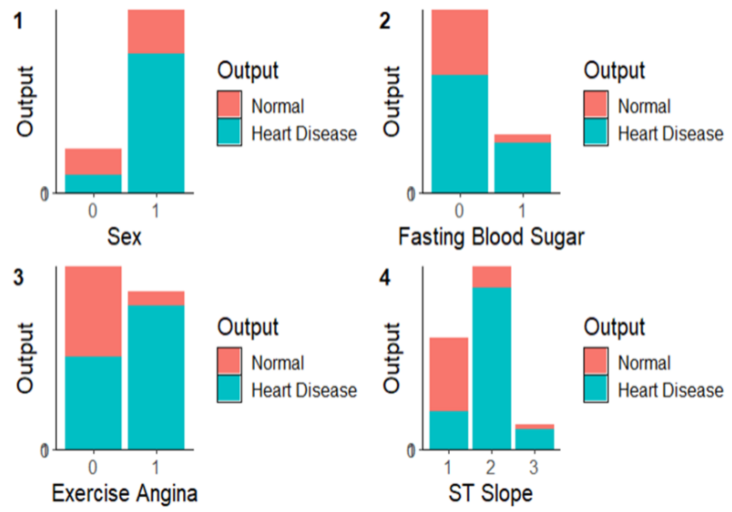

Figure 3. Histogram Showing Relation Between Heart Disease with Several Attributes

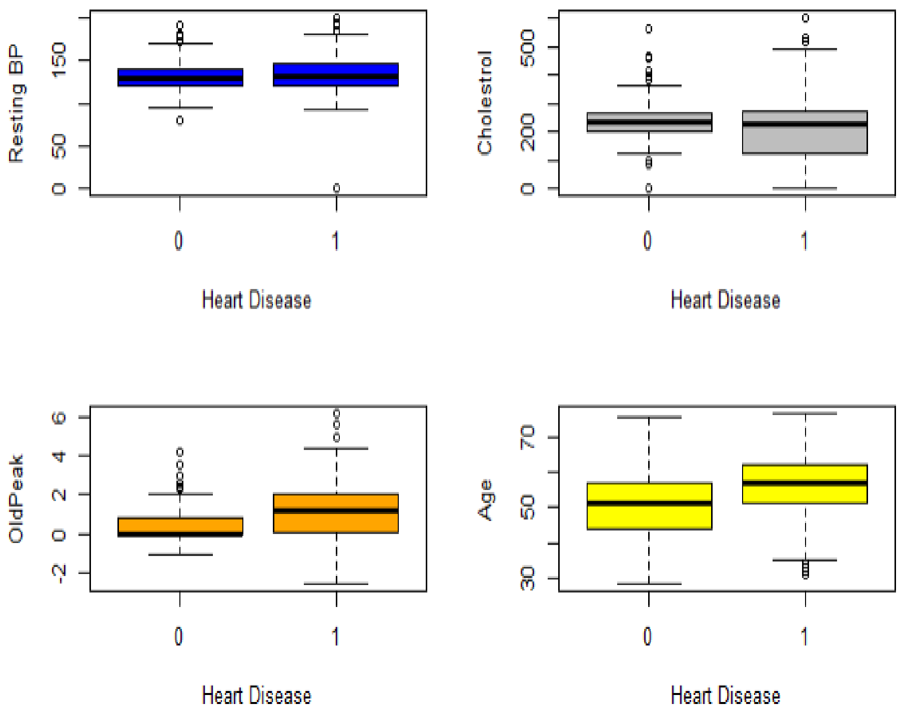

Figure 4. Chi-Squared Plot Showing Relation Between Heart Disease with Several Attributes

Analyses can be performed on each variable classification to learn more about the factors that cause the rising incidence of heart disease in the United States [9]. It is evident that drawing a firm conclusion was challenging because of the asymmetry of the data. However, the analysis showed that males are more likely to develop heart disease than females. Persons with chest pain, ST slope, fasting sugar-related complications, and inactivity were at the highest risk for cardiovascular disease. Researchers also found that smoking was linked to a rise in the incidence of heart disease.

A correlation of 0.5 is considered to be weak. It shows that people of all ages, even pre-teens, can experience heart problems like chest pain. High correlations were found between fasting glucose, age, and the presence or absence of chest pain. The connection between the factors is purely coincidental. Advanced machine learning systems can benefit from the variables. High or low blood sugar levels and diabetes contribute to various forms of heart disease. Cardiovascular disease is most common in the elderly. Figure 2 demonstrates a higher prevalence of heart disease among participants aged 65 and older. This indicates that the prevalence of heart disease increases with age.

We have also tested the dataset with the logistic regression prediction model which is developed using machine learning.

3. TECHNICAL APPROACH

3.1. Inference

The primary objective of the research is to examine the heart disease dataset so that we can predict the window stage of heart disease. Thanks to the prediction, the government will be able to save lives by proactively treating heart disease before it becomes life-threatening.

```
##  Pearson's Chi-squared test

## data:  table(prjdata$chest.pain.type,
prjdata$target)

## X-squared = 334.42, df = 3,
p-value < 2.2e-16
```

Based on the results of the Chi-square test presented above, it is highly unlikely that the two groups are independent. That is why this study rejects the "null hypothesis" explanation. It is safe to say that there's a five percent definitive link between chest pain and the onset of heart disease. The association between specific types of chest pain and the onset of heart disease is thus meaningful. At the same time, it shows evidence linking certain factors to cardiovascular disease at the 5% confidence level.

3.2. The Best Model

Here's a quick rundown of the dataset model: Under residuals, one can see the mean, median, and quartile estimates for each variable.

```
Call:
glm(formula = target ~ sex + chest.pain.type + cholesterol +
    fasting.blood.sugar + exercise.angina + oldpeak + ST.slope,
    family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7193  -0.4575   0.1868   0.5117   2.6788

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -3.088908   0.529001  -5.839 5.25e-09 ***
sex1                  1.557486   0.253180   6.152 7.67e-10 ***
chest.pain.type2     -0.111010   0.434328  -0.256 0.798268
chest.pain.type3     -0.124780   0.391088  -0.319 0.749682
chest.pain.type4      1.651247   0.379469   4.351 1.35e-05 ***
cholesterol          -0.003770   0.001001  -3.767 0.000165 ***
fasting.blood.sugar1  0.842790   0.247969   3.399 0.000677 ***
exercise.angina1      0.933750   0.217272   4.298 1.73e-05 ***
oldpeak               0.513747   0.105803   4.856 1.20e-06 ***
ST.slope2             2.057596   0.215967   9.527  < 2e-16 ***
ST.slope3             0.587023   0.404590   1.451 0.146805
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results from the model analysis shown above can be used to check if the model is adequate for developing the prediction tool. The significance level of the statistical tests was set at 5%. Therefore, we are justified in dismissing types 3, chest pain, type 4, and ST. Slope 2 and type 2 have a strong correlation with cardiac disease. At the 5% significance level, the p-value for the association between chest pain and St is extremely small. The null hypothesis for these tests—exercise angina sex 1, ST.slope 1, old peak, and cholesterol—should also be rejected. The evaluated p-values are greater than the statistical threshold of 0.05. We conclude that these factors play a role in the onset of cardiovascular disease.

3.3. Prediction

Recall, true positive, and hit rate were the tests used to categorize the predictive results as positive. This reveals the conditions that are either 100% present or 100% absent in each sample taken during data collection (true positive or false positive)

```
> # Prediction
> p1 <- predict(mymodel, train, type = 'response')
> head(p1)
         1          2          3          4          5          6
0.06110804 0.21065122 0.06241891 0.82016333 0.08382958 0.05048171
> head(train)
  age sex chest.pain.type resting.bp.s cholesterol fasting.blood.sugar resting.ecg max.heart.rate exercise.angina
1  40   1               2          140         289                   0           0            172               0
2  49   0               3          160         180                   0           0            156               0
3  37   1               2          130         283                   0           1             98               0
4  48   0               4          138         214                   0           0            108               1
5  54   1               3          150         195                   0           0            122               0
6  39   1               3          120         339                   0           0            170               0
  oldpeak ST.slope target
1     0.0        1      0
2     1.0        2      1
3     0.0        1      0
4     1.5        2      1
5     0.0        1      0
6     0.0        1      0
> |
```

Sensitivity 97.5%
Specificity 65.4%

```
         Actual
Predicted   0   1
        0  70   3
        1  37 119
> 1-sum(diag(tab2))/sum(tab2)
[1] 0.1746725
```

The selection process took place in accordance with the correlation outcomes discovered in the data set analysis. Specifically, based on the randomized forest classification method, the Boruta Feature Selection (BFS) algorithm was able to use the same selection procedure. This allowed for the collection of the most fundamental aspects of the dataset.

The test's sensitivity affects identifying the false positives related to cardiovascular disease and other heart diseases [1]. If one is trying to find people who might have heart disease, for instance, and their test results lead them to construct a predictor machine, they can rest assured that the number of false positives will be low. The predictor machine aggregates all available data for a given dataset. After receiving the data, a shadow copy is created. The results from the data set are then communicated to an unnamed classifier per the machine's instructions. The machine inputs the factors that matter most and then uses those factors to rank the rest. If a person is tested and the results come back positive while still valid, then the person is safe. However, the person may experience negative side effects, such as anxiety, as a result. The identified person must have cardiac issues determined by the predictor machine's diagnostic. Therefore, the machine should provide the highest possible risk of having the condition.

3.4. Logistic Regression Model

Logistic regression model is developed using machine learning algorithm. The model is trained using the dataset from IEEE data port which contains 303 rows and 14 column with heart related data.

⇨ Heartdata.describe()

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.313531	0.544554
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.612277	0.498835
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

From the above dataset we can find some insights like 25 percent of the inputs in the dataset are of age 45 and heart disease is observed majorly from the people who are above 45 of age.

⇨ `model = LogisticRegression()`
⇨ `model.fit(A_train, B_train)`
⇨ `A_train_prediction = model.predict(A_train)`
⇨ `training_data_accuracy = accuracy_score(A_train_prediction, B_train)`

Accuracy on Training data : 0.8512396694214877

⇨ `A_Test_prediction = model.predict(A_Test)`
⇨ `testing_data_accuracy = accuracy_score(A_Test_prediction, B_Test)`

Accuracy on Testing data : 0.819

4. DISCUSSION

According to official statistics, nearly half of all Americans have some form of cardiovascular disease. There are several factors at play here. Smoking, unhealthy habits in general, being overweight or obese, getting older, having diabetes, high blood pressure, a family history of these conditions, and not getting enough exercise are all risk factors. Most of the United States' largest healthcare costs have been related to cardiovascular disease. Those who do make it through this ordeal are regular hospital in patients who need a lot of TLC to keep their hearts from failing. Deaths from cardiovascular causes are a major economic setback because they represent a loss of potentially productive minds[2]. In the U.S, coronary heart disease affects both adults and children at a higher rate than any other form of the disease.

The research found that males were likelier to experience heart disease than females. Heart disease can occur in people with fasting blood sugar levels and a heart rate that fluctuates irregularly. The prevalence of heart disease in the U.S has increased as a result of factors including inactivity and smoking. The moderate association of 0.5 suggests that children as young as 12 can experience chest pain due to heart disease. There is a robust correlation between getting older and suffering from chest pain. In this study, participants over 65 had a higher risk of heart disease and a higher mortality rate from cardiovascular causes.

The model of the data set summarizes the results, which helps establish whether or not the model analysis can yield a prediction machine. The p-value for this sample is higher than the statistical threshold. This supports the notion that all the investigated factors are important contributors to the increase of cardiovascular disease. To estimate one's risk of developing heart disease, one must first determine the true positive rate in each sample. The model of the data set's correlation results was then subjected to a filtering process. The method used to cut was based on picking out the most telling features of the dataset.

The predator machine gathers all the information gathered about a person during their testing to create a "shadow copy. "An unnamed classifier is applied to the data, and the highest score is evaluated. The computer determines a person's lifetime risk of developing heart disease and declares the person safe if the results are good and valid.

However, there are restrictions on what can be done to keep heart disease at bay. You cannot make someone younger to keep them from getting heart disease. Decreased immunity is a common side effect of getting older. Reduced immunity makes people more likely to get sick from infections, and heart disease is a real risk, especially for those over 65 who have put on a lot of weight. Physical activity is less effective at preventing heart disease in people of that age. Furthermore, one's genetic makeup cannot be changed. A person's place of birth cannot be altered[7]. This suggests that inheriting a propensity for weight gain is possible. It is also possible that there is a history of early-onset heart disease in their family that could recur.

On the other hand, people should be encouraged to prioritize their health and marry people who come from healthy backgrounds to spread those genes to future generations. At last, gender is immutable. Whether to a male or female, the process of giving birth is natural and cannot be altered. Men need to know why they are more likely to contract heart disease than women.

Estrogen helps women by making them less vulnerable to cardiovascular disease. Keeping this in mind, they ought to take measures to avoid joining the heart disease statistics. All it takes is for them to commit to a healthier way of life.

5. Conclusion

The evaluated p-values are greater than the statistical threshold of 0.05. We conclude that these factors play a role in the onset of cardiovascular disease. the tests indicate that 90 percent of the participant who are facing severe chest pain is suffering from heart disease where majority of the successful heart disease identified is in males and only 10 percent participants are identified as healthy. We have also developed logical regression model based on the machine learning algorithm and we have tested the data using trained data and test data which was developed from dataset. The accuracy of logistic regression model based on training data is 85.12 percent and accuracy of the test data is approximately 82 percent. In conclusion, heart disease symptoms might differ for men and women. Men will probably have chest pains. On the other hand, women tend to have experience different kinds of symptoms together with chest discomfort, for example, shortness of breath, fatigue, and nausea. However, cardiovascular disease can be avoided. It is still necessary to stress the value of prevention over treatment, even with the help of a prediction computer. For cardiovascular health, it is important to check one's blood sugar and cholesterol levels. Regular exercise helps burn calories, reducing the danger of obesity-related heart disease. Heart health is improved, and blood flow is increased due to regular exercise. Eating healthy can help you keep the weight off and protect your heart from the damage caused by atherosclerosis. Informing people of the risks associated with smoking is important because it can lead to serious health problems, including heart disease. Limiting your alcohol intake is another strategy for lowering your risk of developing heart disease. Last but not least, people should learn to cope with stress, as it raises their risk of developing the disease. Stress can precipitate heart disease by increasing blood pressure and should be controlled with exercise or meditation. If people consciously decide to adopt heart-healthy practices, the measures above may be effective in the fight against and reduction of the prevalence of heart disease. Age and cholesterol play significant role heart disease which means higher the value higher the chances of heart diseases.

References

[1] AlGhatrif, M., Cingolani, O., &Lakatta, E. G. (2020). The dilemma of coronavirus disease 2019, aging, and cardiovascular disease: insights from cardiovascular aging science. *JAMA cardiology*, *5*(7), 747-748.https://doi:10.1001/jamacardio.2020.1329

[2] Chobufo, M. D., Singla, A., Rahman, E. U., Michos, E. D., Whelton, P. K., & Balla, S. (2022). Temporal trends in atherosclerotic cardiovascular disease risk among US adults. Analysis of the National Health and Nutrition Examination Survey, 1999–2018. *European Journal of Preventive Cardiology*. https://doi.org/10.1093/eurjpc/zwac161

[3] Chou, R., Cantor, A., Dana, T., Wagner, J., Ahmed, A. Y., Fu, R., & Ferencik, M. (2022). Statin use for the primary prevention of cardiovascular disease in adults: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*, *328*(8), 754-771.https://doi:10.1001/jama.2022.12138

[4] Dinesh K; Nathan S. "Study and Analysis of Chat GPT and its Impact on Different Fields of Study." Volume. 8 Issue. 3, March - 2023 , International Journal of Innovative Science and Research Technology (IJISRT), www.ijisrt.com. ISSN - 2456-2165, PP :- 827-833. https://doi.org/10.5281/zenodo.7767675

[5] Grossman, D. C., Bibbins-Domingo, K., Curry, S. J., Barry, M. J., Davidson, K. W., Doubeni, C. A., ... & US Preventive Services Task Force. (2017). Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults without cardiovascular risk

factors: US Preventive Services Task Force recommendation statement. *Jama*, *318*(2), 167-174.https://doi:10.1001/jama.2017.7171

[6] Kalla, D., & Samaah, F. (2020). *IOSR Journal of Computer Engineering*, *22*(1), 50–56. https://doi.org/10.9790/0661-2201035056

[7] Krist, A. H., Davidson, K. W., Mangione, C. M., Barry, M. J., Cabana, M., Caughey, A. B., ... & US Preventive Services Task Force. (2020). Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: US Preventive Services Task Force recommendation statement. *Jama*, *324*(20), 2069-2075.https://doi:10.1001/jama.2020.21749

[8] Lin, J. S., Evans, C. V., Johnson, E., Redmond, N., Coppola, E. L., & Smith, N. (2018). Nontraditional risk factors in cardiovascular disease risk assessment: updated evidence report and systematic review for the US Preventive Services Task Force. *Jama*, *320*(3), 281-297.https://doi:10.1001/jama.2018.4242

[9] Mangione, C. M., Barry, M. J., Nicholson, W. K., Cabana, M., Chelmow, D., Coker, T. R., ... & US Preventive Services Task Force. (2022). Vitamin, mineral, and multivitamin supplementation to prevent cardiovascular disease and cancer: US Preventive Services Task Force recommendation statement. *JAMA*, *327*(23), 2326-2333.https://doi:10.1001/jama.2022.8970

[10] Mehta, N. K., Abrams, L. R., &Myrskylä, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. *Proceedings of the National Academy of Sciences*, *117*(13), 6998-7000.https://doi.org/10.1073/pnas.1920391117

[11] ‘O’Connor, E. A., Evans, C. V., Ivlev, I., Rushkin, M. C., Thomas, R. G., Martin, A., & Lin, J. S. (2022). Vitamin and mineral supplements for the primary prevention of cardiovascular disease and cancer: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*, *327*(23), 2334-2347.https://doi:10.1001/jama.2021.15650

[12] Patnode, C. D., Evans, C. V., Senger, C. A., Redmond, N., & Lin, J. S. (2017). Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US Preventive Services Task Force. *Jama*, *318*(2), 175-193.https://doi:10.1001/jama.2017.3303

[13] Kuraku, S.; Kalla, D. Emotet Malware–A Banking Credentials Stealer. Iosr J. Comput. Eng. 2020, 22, 31–41.

[14] Sturgeon, K. M., Deng, L., Bluethmann, S. M., Zhou, S., Trifiletti, D. M., Jiang, C., ... &Zaorsky, N. G. (2019). A population-based study of cardiovascular disease mortality risk in US cancer patients. *European heart journal*, *40*(48), 3889-3897.https://doi.org/10.1093/eurheartj/ehz766

[15] Tobias, D. K., Stuart, J. J., Li, S., Chavarro, J., Rimm, E. B., Rich-Edwards, J., ... & Zhang, C. (2017). Association of history of gestational diabetes with long-term cardiovascular disease risk in a large prospective cohort of US women. *JAMA internal medicine*, *177*(12), 1735-1742.https://doi:10.1001/jamainternmed.2017.2790

[16] Xu, G., Snetselaar, L. G., Strathearn, L., Ryckman, K., Nothwehr, F., & Torner, J. (2022). Association between history of attention-deficit/hyperactivity disorder diagnosis and cardiovascular disease in US adults. *Health Psychology*.https://psycnet.apa.org/doi/10.1037/hea0001193

[17] Zhou, D., Xi, B., Zhao, M., Wang, L., &Veeranki, S. P. (2018). Uncontrolled hypertension increases risk of all-cause and cardiovascular disease mortality in US adults: the NHANES III Linked Mortality Study. *Scientific reports*, *8*(1), 1-7. https://doi:10.1001/jama.2022.12138

AUTHORS

Dinesh Kalla is currently working at Microsoft as Big Data and Azure Cloud Escalation Engineer and has 8 years of industry experience as a .Net Developer, BI Developer and Azure Cloud Engineer. His main areas of expertise and research interest are in Big Data Analytics, Data Science, Machine Learning, Artificial Intelligence, IOT and Cybersecurity. He published several papers related to Chatbots and cybersecurity threats in international Journals. He completed his Masters in University of New Haven and currently pursuing his Doctoral Degree in Computer Science specialized in Big Data from Colorado Technical University.

Arvind Chandrasekaran from Texas, USA. Presently working for PPG Healthcare for the past six years. I have also been pursuing Doctorate in Computer Science (Big Data Analytics) from Colorado Technical University for the past two years, having completed 54 credits; I'm looking to achieve the same by this year.