CARDIAC ARREST PREDICTION USING MACHINE LEARNING

Srimedha Bhavani Chandoo

Jagrati Chauhan

Sai Pratheek KVDSNK

Navya Prasad Malur Narasimha Prasad

Sai Krishna Sriram

December 19, 2023

Abstract

The cardiovascular disease known as coronary heart disease (CHD) is characterized by narrowing or blockage of the blood arteries that nourish the heart muscle. This results in decreased blood flow and possible heart-related problems. If left untreated, this may possibly result in a heart attack. To implement a Machine Learning Model that can accurately predict cardiac arrest risk based on readily available patient data is the primary goal of the project. Finding the elements that have the biggest effects on a patient is another aspect of this study. After the data has been cleaned and processed, understanding the influence of various variables is being done utilizing the many visualization approaches available. Following the completion of data balancing and feature selection, five distinct machine learning models — logistic regression, random forest, gradient boost model, k-nearest neighbors algorithm and support vector machine — are put into practice.

1 Introduction

Cardiac arrest, a sudden and life-threatening event, continues to pose a significant challenge in global healthcare. Despite advances in medical science, accurate and timely

prediction of cardiac arrest remains elusive. In this study, we leverage machine learning techniques to develop a predictive model, aiming to enhance the precision of cardiac arrest prediction. The need for reliable prediction tools is underscored by the potential to intervene early, saving lives and alleviating the strain on healthcare resources. Through innovative approaches and comprehensive data analysis, this research seeks to contribute to the advancement of cardiac arrest prediction, addressing the limitations of current methodologies.

Because of the global effect of Coronary Heart Disease (CHD), identifying common risk factors is critical. CHD is a primary cause of morbidity and mortality, and knowing these risk factors has important public health consequences. It enables tailored interventions, which improves the efficiency of healthcare resource allocation. Individually, understanding enables people to make more educated lifestyle choices, which leads to better health outcomes. This research promotes innovation by leading breakthroughs in cardiovascular health and tailored healthcare solutions. Aside from individual health, addressing risk factors aids in the reduction of health disparities and promotes equity. The social and economic impact of CHD emphasizes the necessity of preventative efforts, which present a potential to reduce both individual and community burdens linked with the illness.

As mentioned in the abstract we selected three different machine learning models to predict Coronary Heart Disease. Random Forest is an ensemble learning method that builds multiple decision trees and merges their predictions to enhance accuracy, Support Vector Machine is a supervised learning algorithm that finds an optimal hyperplane to classify data points, and Logistic Regression is a statistical method for binary classification predicting the probability of an event occurring. After the models have been implemented we are planning to perform hypothesis testing on two different hypotheses. Hypothesis testing is a statistical method used to assess the validity of a claim or hypothesis about a population parameter by analyzing sample data [2].

2 Data

Our research makes use of NHANES data from 1999-2000 to 2015-2016. The dataset is created by merging the demographic, examination, laboratory, and questionnaire data of 37,079 (CHD - 1300, Non-CHD - 35,779) persons. The age and gender of survey participants at the time of screening are demographic characteristics. The examination data on participant weight, height, blood pressure, and body mass index (BMI) are also treated as a collection of risk factor variables to examine their effect on cardiovascular illnesses. Depending on their age and gender, NHANES collects laboratory and survey data from individuals every two years. Furthermore, a thorough list of risk factor factors is picked from laboratory tests conducted based on previously validated experimental research. Questionnaire data consists of questions asked at home by interviewers utilizing the NHANES website's Computer-Assisted Personal Interview (CAPI) system [3] [1].

3 Methods

3.1 Data Pre-processing

Initially, an attempt was made to obtain an overall grasp of the data set, including a comprehension of the data types included and an overall summary statistics of the data. The presence of null values and outliers was checked, and fortunately, the data was already pretty clean, with no null values or outliers. Outliers were checked for using the method of Z-Score calculation with a threshold of 2.

Even though the data was free of outliers and null values, we encountered a significant data imbalance problem. To address this issue, we used a combination of over and under-sampling of the data to ensure that no data was duplicated and no data points were lost. We utilized the standard under-sampling technique and SMOTE for over-sampling. This is discussed in detail in the next subsection.

3.2 Data Balancing

Techniques for dealing with imbalanced classes include oversampling and under-sampling, Over-sampling duplicates minority class instances, and balancing class distribution. Under-

Column Name	Null Count	Null Percentage
SEQN	0	0.00%
Gender	0	0.00%
Age	0	0.00%
Annual-Family-Income	0	0.00%
Ratio-Family-Income-Poverty	0	0.00%
X60-sec-pulse	0	0.00%
Systolic	0	0.00%
Diastolic	0	0.00%
Weight	0	0.00%
Height	0	0.00%
Body-Mass-Index	0	0.00%
White-Blood-Cells	0	0.00%
Lymphocyte	0	0.00%
Monocyte	0	0.00%
Eosinophils	0	0.00%
Basophils	0	0.00%
Red-Blood-Cells	0	0.00%
Hemoglobin	0	0.00%
Mean-Cell-Vol	0	0.00%
Mean-Cell-Hgb-Conc.	0	0.00%
Mean-cell-Hemoglobin	0	0.00%
Platelet-count	0	0.00%
Mean-Platelet-Vol	0	0.00%
Segmented-Neutrophils	0	0.00%
Hematocrit	0	0.00%
Red-Cell-Distribution-Width	0	0.00%
Albumin	0	0.00%
ALP	0	0.00%
AST	0	0.00%
ALT	0	0.00%
Cholesterol	0	0.00%
Creatinine	0	0.00%
Glucose	0	0.00%
GGT	0	0.00%
Iron	0	0.00%
LDH	0	0.00%
Phosphorus	0	0.00%
Bilirubin	0	0.00%
Protein	0	0.00%
Uric.Acid	0	0.00%
Triglycerides	0	0.00%
Total-Cholesterol	0	0.00%
HDL	0	0.00%
Glycohemoglobin	0	0.00%
Vigorous-work	0	0.00%
Moderate-work	0	0.00%
Health-Insurance	0	0.00%
Diabetes	0	0.00%
Blood-Rel-Diabetes	0	0.00%
Blood-Rel-Stroke	0	0.00%
CoronaryHeartDisease	0	0.00%

Figure 1: Null Value count and percentage for each feature

```
Before Data Balencing
 CoronaryHeartDisease
0    35571
1     1508
Name: count, dtype: int64
```

```
After Data Balencing
 CoronaryHeartDisease
0    15080
1    15080
Name: count, dtype: int64
```

Figure 2: Count of Data points with and without Coronary Heart Disease (Imbalenced vs Balenced data)

sampling involves the removal of data points from the majority class. SMOTE (Synthetic Minority Over-sampling Technique) which is one of the most commonly used oversampling techniques creates synthetic minority class instances by interpolating existent ones. Over-sampling minimizes information loss, while under-sampling reduces computing demands; SMOTE solves both, improving model performance in instances where class imbalances affect prediction accuracy. Each approach involves trade-offs, and their choice is determined by the qualities and purposes of the data.

3.3 Feature Selection and Extraction

After balancing the data, we are left with a dataset with 50 features and 1 target column. As we can't fully utilise all of the columns for our analysis, we tried to reduce the number

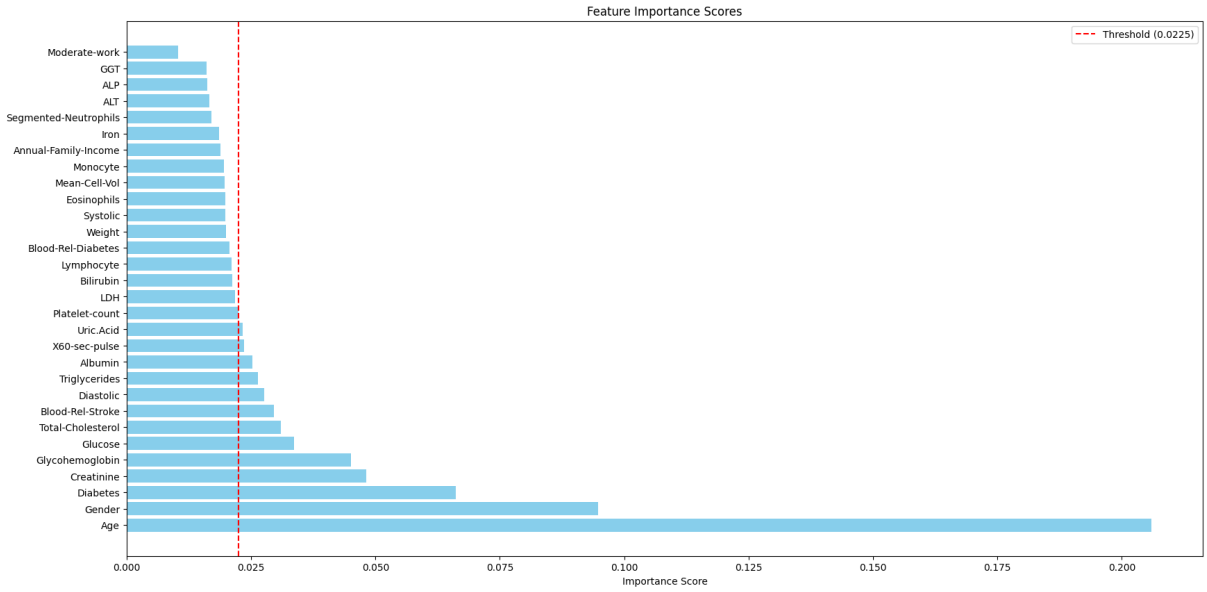


Figure 3: Feature Importance scores assigned by the Random Forest algorithm

of features and retain the most significant columns which affect the target variable. We used the following feature selection methods,

1. **Chi-Square test:** We utilized the Chi-Square test to select 30 features from the initial set of 49 characteristics. The chi-squared test is a statistical method used to determine if there is a significant association between categorical variables. It compares observed and expected frequencies, assessing whether deviations are beyond what might occur by chance, aiding in hypothesis testing for independence in contingency tables.

2. **Random Forest feature importance score:** We used the features obtained from the previous step to train a Random Forest model, to determine the relevance of each feature. The Random forest model uses Gini importance method to compute the feature importances by varying the mean impurities of each tree in the forest. After obtaining feature importances, we have shortlisted them based on a threshold value such that we select only the features which affect our target variable CoronaryHeartDisease the most.

Based on these importance scores, we eventually chose thirteen features for further analysis and hypothesis testing. The final feature set comprises the following features,

- Age
- Gender
- Diabeties
- Creatinine
- Glycohemoglobin
- Glucose
- Total-Cholesterol
- Blood-Rel-Stroke
- Diastolic
- Triglycerides
- Albumin
- X60-sec-pulse
- Uric.Acid

3.4 Exploratory Data Analysis

The graphical depiction of information allows for easy comprehension. Through visual aspects, it improves understanding, finding trends, and helps in the efficient communication of complicated data sets. We used data visualization to gain a better grasp of the relationships between variables and to spot any general trends and patterns that might exist in the dataset.

3.4.1 Histogram of Age & Count

Observations :

The incidence of CHD rises with advancing age - The graph clearly illustrates an pattern indicating that the number of individuals diagnosed with CHD increases as they grow older. This aligns, with known risk factors for CHD, where age plays a role.

Increase among older individuals - The upward trend seems to intensify after approximately the age of 50 - 60. This suggests that the risk of developing CHD escalates rapidly, in age groups.

Individual variations observed - While the graph depicts a trend there is also individual diversity. Some individuals develop CHD at ages while others maintain health well into their later years. This underscores the importance of considering individual risk factors beyond considering age.

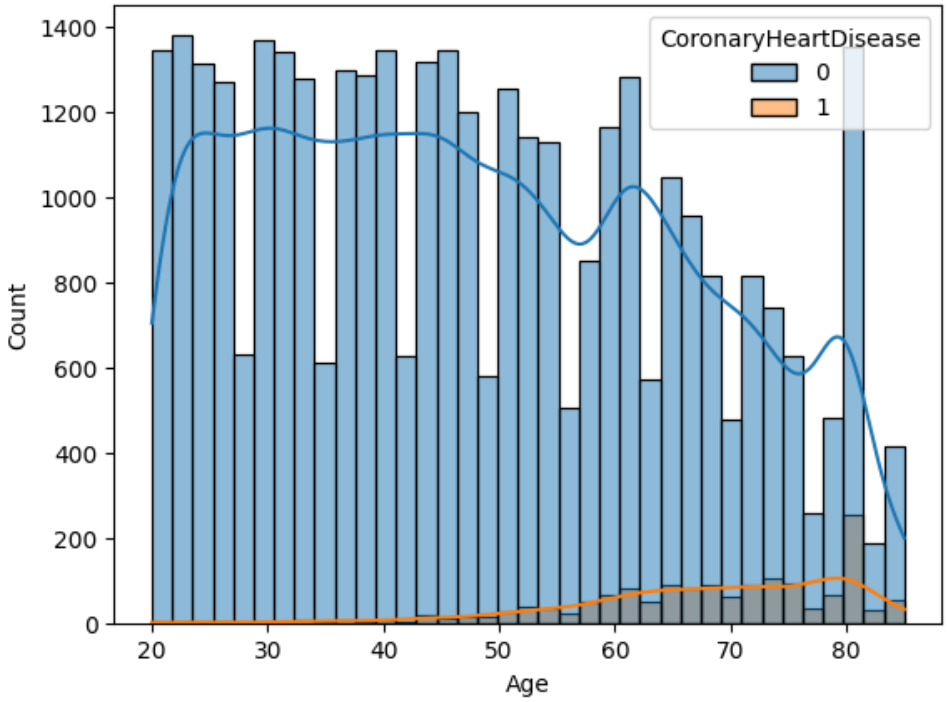


Figure 4: Relationship between age and the number of people diagnosed with & without Coronary Heart Disease

3.4.2 Gender Distribution using Pie Chart

Observations :

Men are more susceptible, to experiencing Coronary Heart Disease (CHD) compared to women. This statistical pattern suggests that lifestyle choices may play a role in this discrepancy. One of the contributing factors to this gap is the incidence of specific risk factors among men. For instance men are more likely to have cholesterol levels due to habits like consuming food and smoking both of which are associated with CHD. Elevated cholesterol can lead to the hardening of arteries thereby increasing the likelihood of heart attacks and strokes. On the hand tobacco use damages blood. Raises the risk of heart disease.

These risk factors are more prevalent in men resulting in a frequency of CHD cases. However it is essential to note that while these types of risk factors are primarily identified in males they can also affect females as they are not immune, to CHD. Nonetheless men generally have a number of risk factors and a higher incidence rate of CHD. To effectively manage these risks and prevent CHD both genders should adopt lifestyle behaviors. Undergo regular medical check ups.

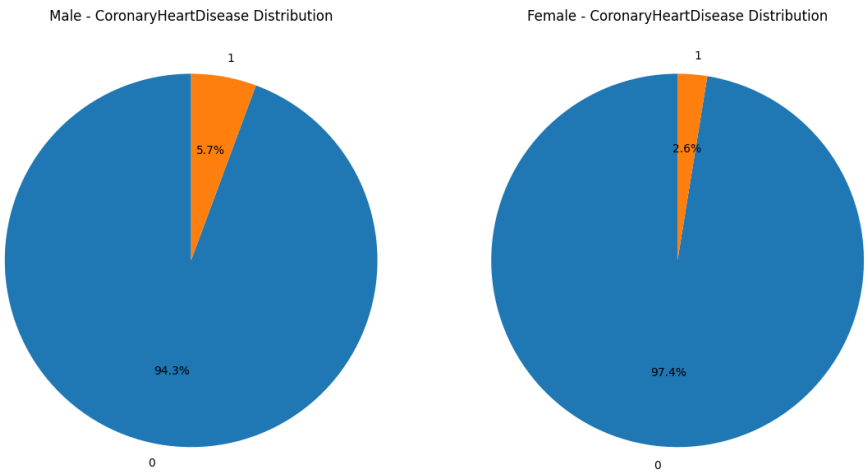


Figure 5: Distribution between males and females diagnosed with & without Coronary Heart Disease

3.4.3 Uric Acid Levels using Bubble Plot

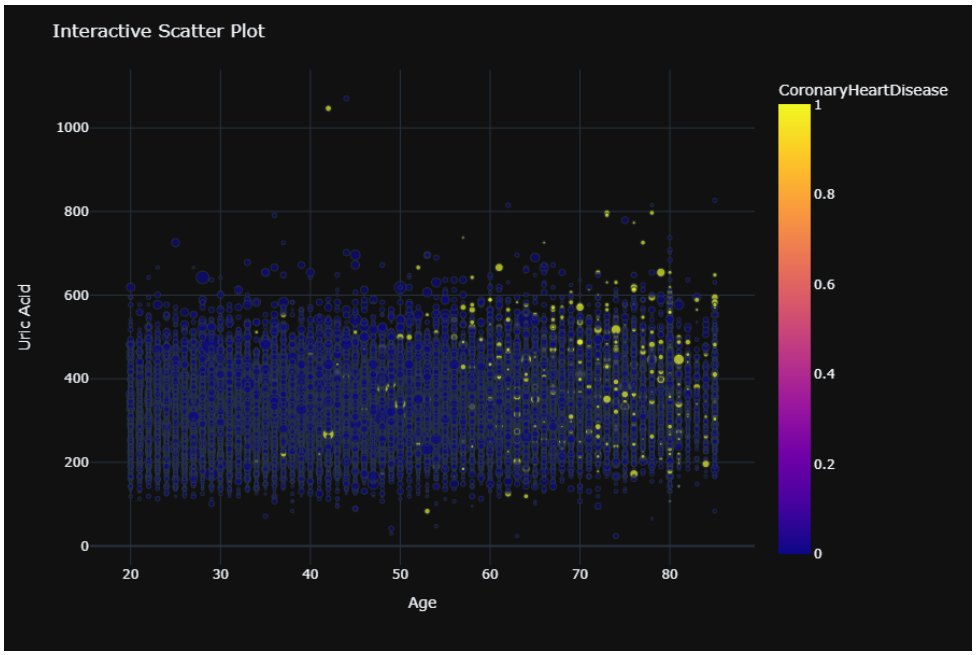


Figure 6: Determining levels of Uric Acid of people by Age diagnosed with & without Coronary Heart Disease

Observations :

As the age is increasing, the people who are diagnosed with Cardiac Heart Disease, their Uric Acid Levels are spiking up. As we get older our bodies naturally start producing triglycerides and uric acid. While these compounds are essential, but excessive levels can pose health risks. Triglycerides are a kind of fats that might accumulate in our blood vessels which in turn contribute to the formation of plaque, known as atherosclerosis,

which leads to some heart disease eventually, and in turn might result in heart stroke. On the hand uric acid is a waste product that can become harmful when it builds up in our bloodstream. The high levels of this acid might develop gout, which causes a very painful inflammation in the joints. Also increase the risk of coronary heart disease. Therefore as we age there is an increased chance for our bodies to produce amounts of these substances indirectly raising the risk of developing heart disease and experiencing arrest. That's why it's vital to maintain a lifestyle and regularly monitor these substances to prevent health conditions.

3.4.4 Line Plot for Cholesterol Levels Distribution

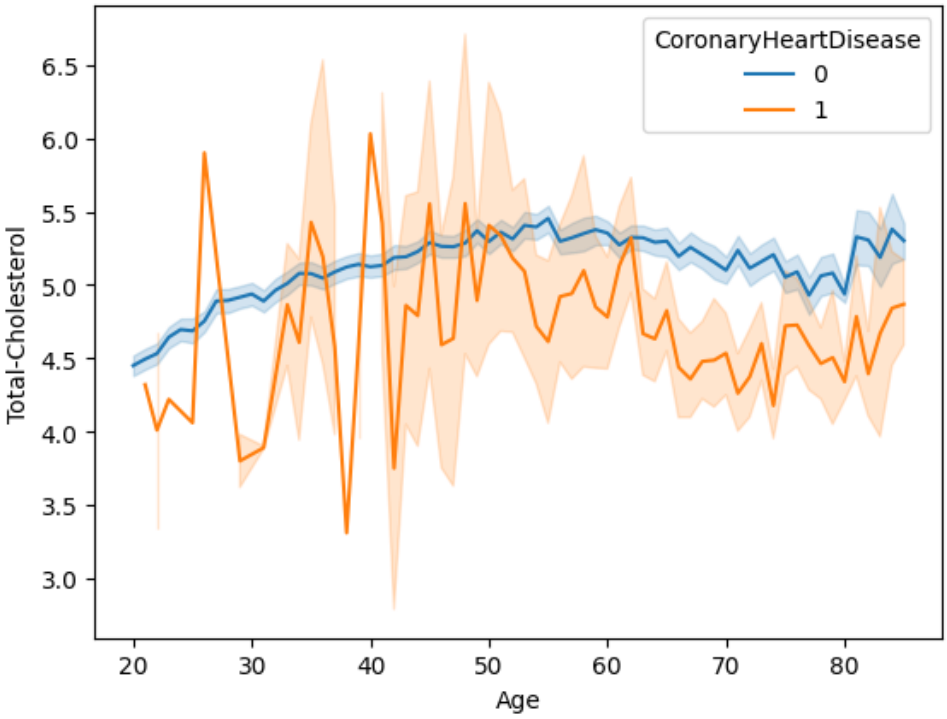


Figure 7: Total Cholesterol levels distribution of people by Age diagnosed with & without Coronary Heart Disease

Observations :

Individuals with higher or lower cholesterol levels than the recommended range tend to have a greater potential for developing CHD. People with consistently normal cholesterol levels within the healthy range generally face a significantly lower risk of CHD. Cholesterol is a type of fatty substance which moves through our bloodstream. It plays an important role in our overall well being. It is usually transported by some proteins known as "Lipoproteins" (Those are LDL - low density lipoprotein and HDL - high density

lipoprotein, known as bad and good cholesterols respectively). So, when the cholesterol levels makes deviation from the acceptable range, the risk of getting CHD to individuals with high LDL cholesterol levels is more because of plaque formation in the arteries, this condition is known as atherosclerosis. This can further lead to have heart attacks and strokes. The individuals who constantly try to maintain their cholesterol levels within the prescribed range i.e., with high HDL levels, they have a very less chance of getting CHD. This is because HDL cholesterol carries the cholesterol from the blood stream to the liver for elimination. Hence, it is very crucial for individual's heart health to keep their cholesterol levels within the prescribed range, by doing regular check-ups and following a healthy lifestyle.

3.4.5 Relationship between Creatinine & Uric Acid using Scatter Plot

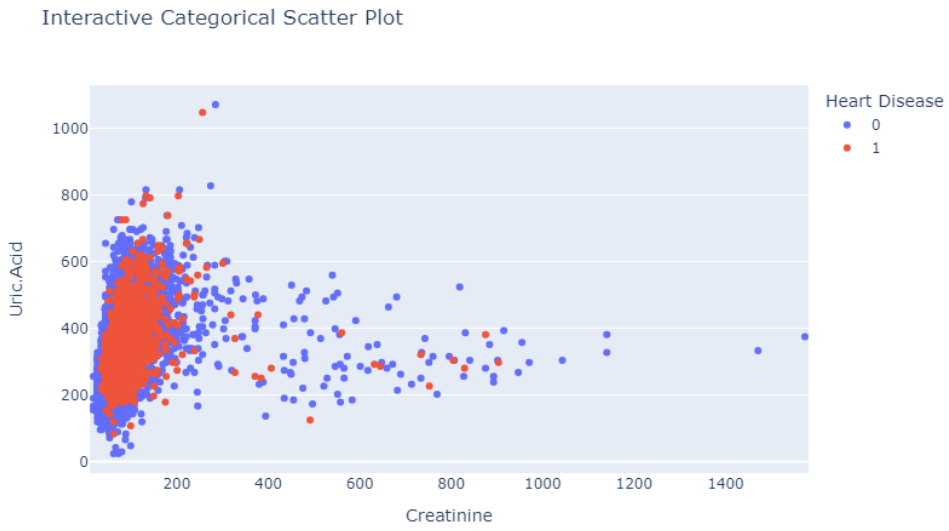


Figure 8: Relationship between Creatinine & Uric Acid of people diagnosed with & without Coronary Heart Disease

Observations :

The relationship between creatinine and uric acid seems complex and non-linear. While there's a general trend of uric acid increasing with creatinine, individual data points deviate from this trend, highlighting the influence of other factors. For suppose, the demographic characteristics like sex, age, ethnicity can influence both creatinine and uric acid levels. And some factors like diet, exercise and smoking can also play a huge rule. Consider an example, a diet high in purines which are building blocks of uric acid,

might lead to higher uric acid levels. On the other side, regular workout can help in regulating the uric acid levels by creatinine's production. Furthermore, some of the medical conditions can also affect these levels, like Chronic Kidney Disease (CKD) could lead to higher ratio of uric acid and creatinine results in high risk of CKD. Although there is some pattern in rise of uric acid with creatinine, the relationship can be changed by different circumstances, making it a very complicated relationship. Understanding these parameters can aid in the creation of more tailored and effective strategies for uric acid management and the prevention of illnesses like gout and kidney disease.

3.5 Hypotheses Testing

Hypothesis testing is a statistical strategy for drawing conclusions based on a sample of data. Hypothesis testing can be used to validate the machine learning model's predictions.

Hypothesis testing involves the following steps:

- Formulating the Hypothesis: The initial stage in hypothesis testing is to develop the null hypothesis (H_0) and the alternative hypothesis (H_1). The null hypothesis is the initial claim based on popular belief. The alternative hypothesis is the assertion that we want to prove to be true.

- Collecting Data: After formulating the hypotheses, data is collected to test the hypotheses. The data collection should be designed to test the hypotheses.

- Selecting the Type of Hypothesis Test: The type of statistical test to use can be determined by predictor variable, it can be any of quantitative or categorical variables. Tests like Chi-Square test is used for categorical data and Z-Test, T-Test, F-Test are generally used in case of quantitative data.

- Deciding whether to Reject or Fail to Reject the Null Hypothesis: Based on the outcome of the statistical test obtained, a decision can be made whether to reject or fail to reject the null hypothesis. Also, the p-value which is generated by that statistical test is used for further guidance of decision.

- Presenting the Findings: The initial hypothesis is discussed in terms of whether it was supported by the results or not.

3.5.1 Hypothesis 1: T-Test for Age

Null Hypothesis (H_0): The mean age of individuals with Coronary Heart Disease is equal to the mean age of individuals without Coronary Heart Disease

Alternative Hypothesis (H_A): The mean age of individuals with Coronary Heart Disease is not equal to the mean age of individuals without Coronary Heart Disease

Result: Null Hypothesis is rejected.

$$H_0 : \mu_{\text{age\_with\_CHD}} = \mu_{\text{age\_without\_CHD}}$$

$$H_1 : \mu_{\text{age\_with\_CHD}} \neq \mu_{\text{age\_without\_CHD}}$$

An independent samples t-test was performed on the age distributions of the two groups to test this hypothesis. The t-statistic and p-value were computed, with 0.05 chosen as the significance level (α).

The t-test findings revealed that the p-value ($p_{\text{value\_age}}$) was smaller than α. As a result, the null hypothesis is rejected, showing that there is a substantial difference in mean age between people with and without Coronary Heart Disease.

As a result of this hypothesis, we can conclude that age is a key factor in prediction of Coronary Heart Disease in humans

3.5.2 Hypothesis 2: T-Test for Glycohemoglobin

Null Hypothesis (H_0): The mean Glycohemoglobin level is equal in individuals with and without Coronary Heart Disease

Alternative Hypothesis (H_A): The mean Glycohemoglobin level is not equal in individuals with and without Coronary Heart Disease

Result: Null Hypothesis is rejected.

$$H_0 : \mu_{\text{glycohemoglobin\_with\_CHD}} = \mu_{\text{glycohemoglobin\_without\_CHD}}$$

$$H_1 : \mu_{\text{glycohemoglobin\_with\_CHD}} \neq \mu_{\text{glycohemoglobin\_without\_CHD}}$$

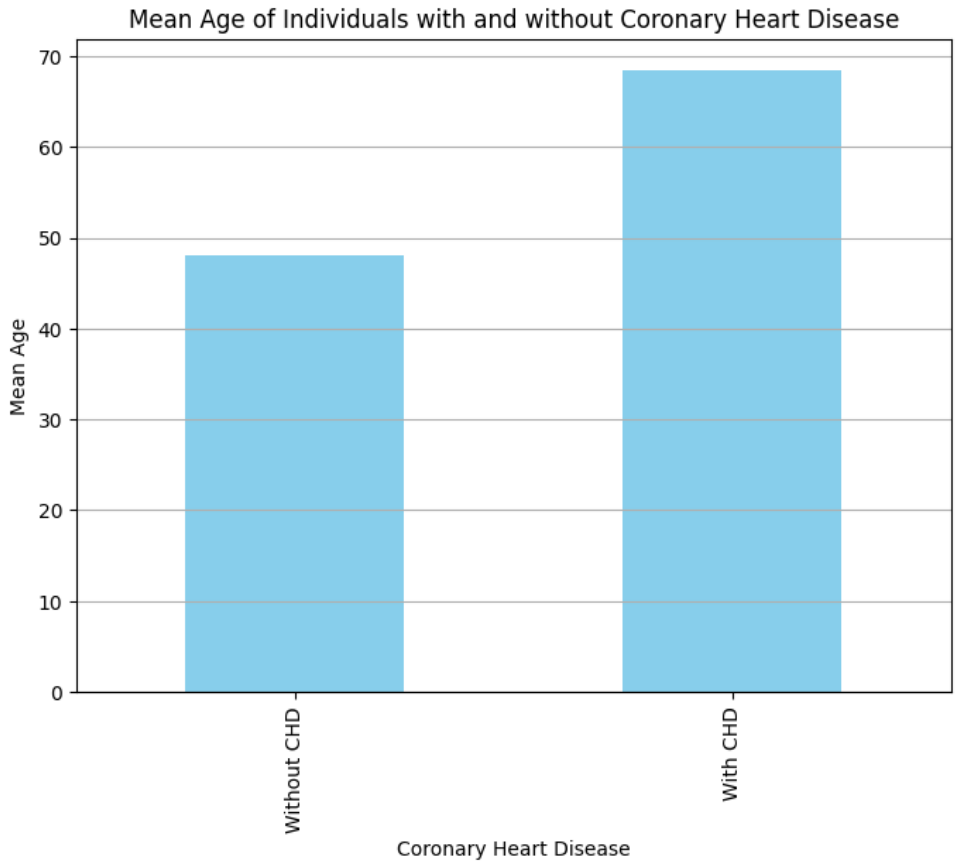


Figure 9: Bar Graph depicting the mean age of individuals with and without Coronary Heart Disease

The second hypothesis aims to find the effect of Glycohemoglobin levels in patients with Coronary Heart Disease. An independent samples t-test was performed on the glycohemoglobin level distributions of the two groups to test this hypothesis. The t-statistic and p-value were computed, with 0.05 chosen as the significance level (α).

The t-test findings revealed that the p-value ($p_{\text{value\_age}}$) was smaller than α. As a result, the null hypothesis is rejected, showing that there is a substantial difference in mean glycohemoglobin levels between people with and without Coronary Heart Disease.

This means that glycohemoglobin level is an important feature in predicting Coronary Heart Disease in patients.

3.5.3 Hypothesis 3: Mann-Whitney U Test for Pulse Rate

Null Hypothesis (H_0): Having a history of Coronary Heart disease does not have any effect on the pulse rate of people

Alternative Hypothesis (H_A): Having a history of Coronary Heart disease has an ef-

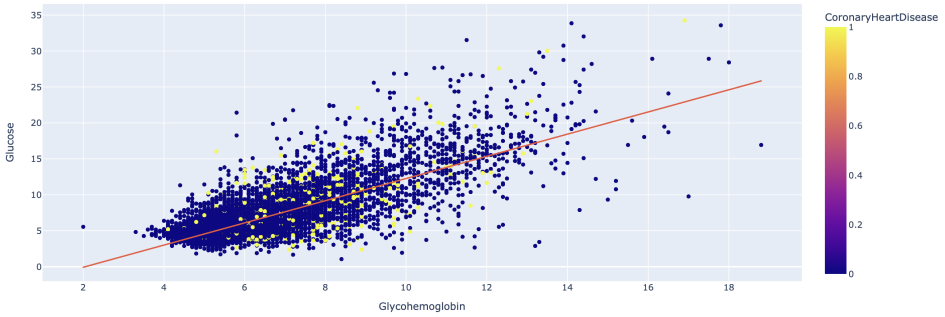


Figure 10: Scatter plot with thread line depicting the mean glycohemoglobin levels of individuals with and without Coronary Heart Disease

fect on the pulse rate of people

Result: Null Hypothesis is rejected.

The third hypothesis aims to find the relation between pulse rate of individuals and their history of blood related stroke. The test was performed, and the results indicated a Mann-Whitney U Statistic of 115365137.5 and a p-value ($p_{\text{value\_pulse}}$) of $4.452551164503746 \times 10^{-6}$. The p-value being less than α leads to the rejection of the null hypothesis, suggesting that having a history of Coronary Heart Disease has a significant effect on the pulse rate of individuals.

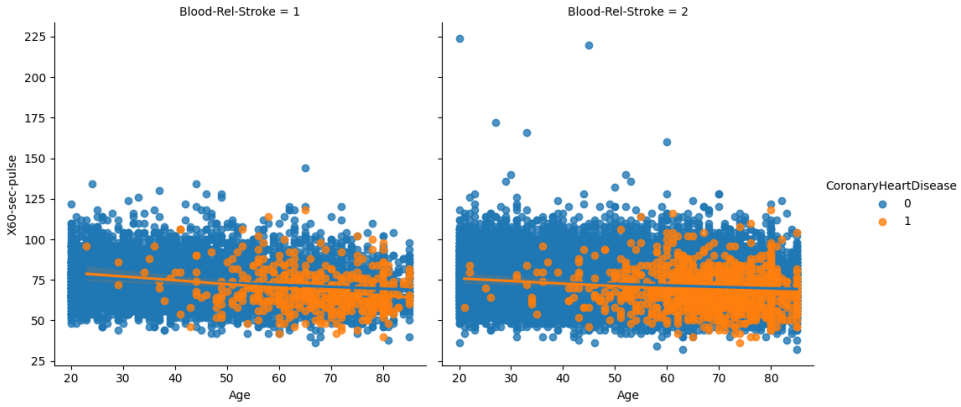


Figure 11: Scatter plots with overlaid regression lines

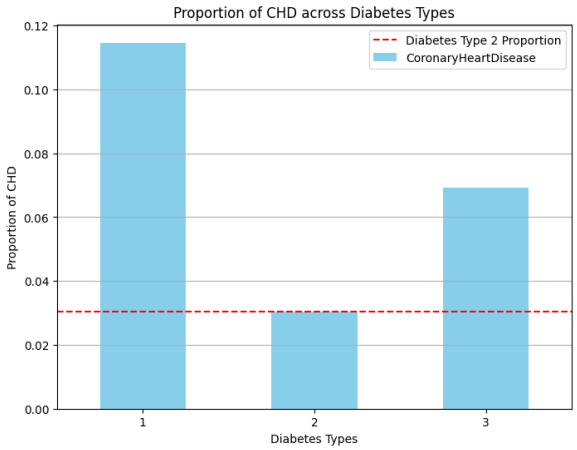
3.5.4 Hypothesis 4: Chi2 Test for Diabetes

Null Hypothesis (H_0): The proportion of people with Coronary Heart Disease is the same across all diabetes types

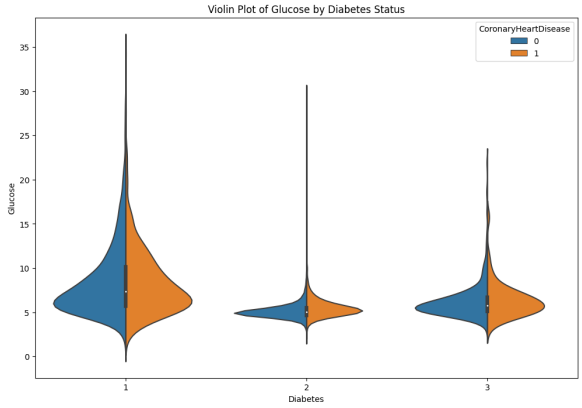
Alternative Hypothesis (H_A): The proportion of people with Coronary Heart Disease is higher among Type 2 Diabetes patients compared of other types of diabetes

Result: Null Hypothesis is rejected.

The fourth hypothesis aims to find the association between Coronary Heart Disease and different types of diabetes types using Chi-squared test. A contingency table was created, and the Chi-squared test was performed. The results yielded a Chi-square test statistic and a p-value ($p_{\text{value\_diabetes}}$) of $1.665500066313612 \times 10^{-148}$. With the p-value less than α, the null hypothesis is rejected, providing evidence that the proportion of people with Coronary Heart Disease differs across the 3 different diabetes types. From the graphs we can see that Coronary Heart Disease is higher among individuals with Type 1 Diabetes, followed by Type 3 and Type 2.



(a) Proportions of CHD occurrences for each type of Diabetes



(b) Violin Plot of Glucose by Diabetes Status

Figure 12: Hypotheses 4

3.5.5 Hypothesis 5: T-Test for Pulse Rate and Age

Null Hypothesis (H_0): The pulse rate varies the same for people with and without Coronary Heart Diseases

Alternative Hypothesis (H_A): The pulse rate varies rigorously for people with Coronary Heart Disease than people without any heart disease

Result: Null Hypothesis is rejected.

The fifth hypothesis aims to find the association of variation in pulse rate w.r.t age, between individuals with and without Coronary Heart Disease effect of age in Coronary Heart Disease. A t-test was conducted to assess whether the pulse rate varies more rigorously for people with CHD than those without. The results yielded a t-statistic and a p-value ($p_{\text{value\_pulse\_age}}$) of $2.155744320482476 \times 10^{-41}$. With the p-value less than α, the null hypothesis is rejected, suggesting that there is evidence to suggest that the pulse rate varies more rigorously for people with CHD than those without.

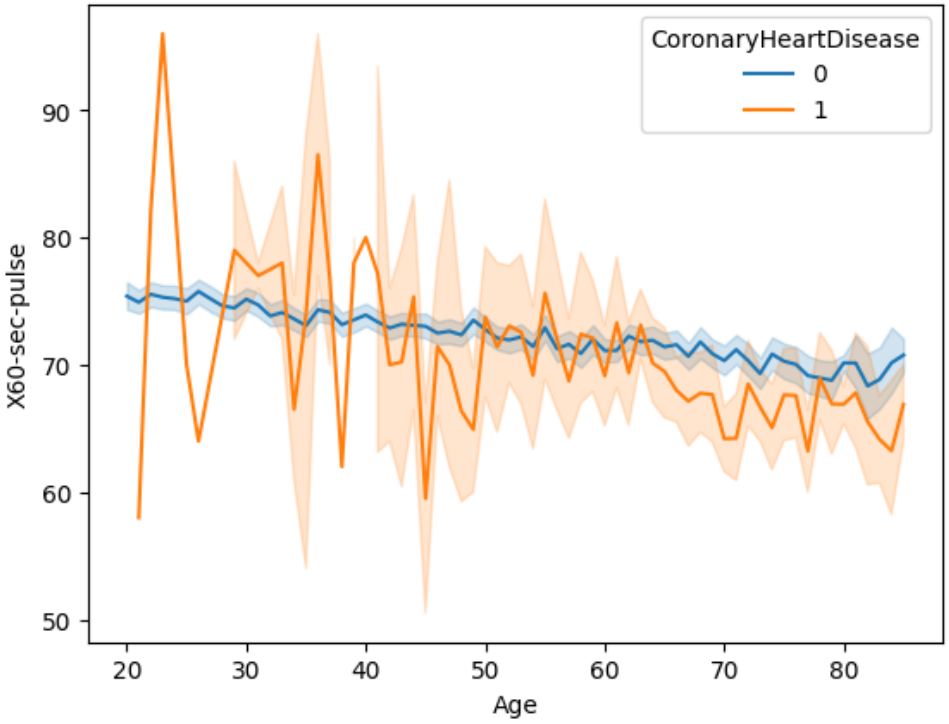


Figure 13: Line Plot of Age vs X60 pulse

3.6 Model Training & Evaluation

In the data set used, we are focusing on 'CoronaryHeartDisease' as the target variable: Does the patient have the disease or not? It is a yes (1) or no (0) situation. To make predictions, we're employing five different models – logistic regression, random forest, gradient boost model, k-nearest neighbors algorithm and support vector machine – each trying to tell us whether someone has heart disease or not. It is similar to having different experts giving their insights on whether a person might be at risk.

Support Vector Machine : SVM is a versatile supervised learning technique used for classification and regression. It locates a hyperplane in a three-dimensional space

while maximizing the margin between classes. SVM tackles nonlinear relationships by using the kernel approach and indirectly altering data. Support vectors, which are important data points that define the margin, add to its resilience. The 'C' parameter strikes a balance between classification accuracy and margin size. Text categorization, picture recognition, and bioinformatics all use SVM. SVM's computational intensity and sensitivity to kernel and parameters need careful implementation in high-dimensional spaces. It excels in handling large datasets and making accurate predictions.

Random Forest : Random Forest is an ensemble learning technique that, during training, creates a large number of decision trees and then merges their predictions to increase accuracy and robustness. It chooses subsets of characteristics at random for each tree, increasing diversity and decreasing overfitting. The system performs well in classification and regression tasks, making accurate predictions on large datasets. It reduces the possibility of individual tree faults and manages missing data. Random Forest is widely used in a variety of sectors, including banking, healthcare, and remote sensing, due to its ease of use, scalability, and ability to capture subtle correlations. It has demonstrated great performance and durability against noisy data.

Logistic Regression : Logistic Regression is a statistical method for classifying binary and multiclass data. It is used for classification rather than regression, despite its name. The logistic function is applied to a linear combination of input features to model the probability of a binary result. The method learns coefficients using optimization approaches, with a focus on simplicity and interpretability. To avoid overfitting, regularization techniques such as L1 and L2 might be used. Logistic Regression is frequently utilized in sectors such as medical, finance, and social sciences because of its efficiency, simplicity, and ease of interpretation in generating probabilities and assisting decision-making in situations where understanding the likelihood of events is critical.

Gradient Boosting : Gradient Boosting is a stage-wise ensemble learning technique that combines the capabilities of numerous weak learners, typically decision trees, to produce a predictive model. It reduces mistakes by fitting new models sequentially

to the residual errors of preceding ones. Popular implementations include AdaBoost and XGBoost. Each tree corrects the mistakes of its forefathers, resulting in a robust and accurate model. Gradient Boosting performs well in regression and classification problems, can handle complex relationships, and has a high predictive accuracy. However, careful tuning is required, and it may be computationally demanding, making it appropriate for a variety of machine-learning applications.

K-Nearest Neighbors : KNN is a straightforward and adaptable supervised machine-learning technique that may be utilized for classification and regression applications. It classifies a data point in the feature space by evaluating the majority class of its k nearest neighbors. The value of k influences how sensitive the model is to local fluctuations. Because it does not explicitly learn a model during training, KNN is non-parametric and lazy. While KNN is successful for small to moderately-sized datasets, it may incur substantial processing costs when dealing with huge datasets. Proper k selection and feature scaling consideration are critical for optimal performance.

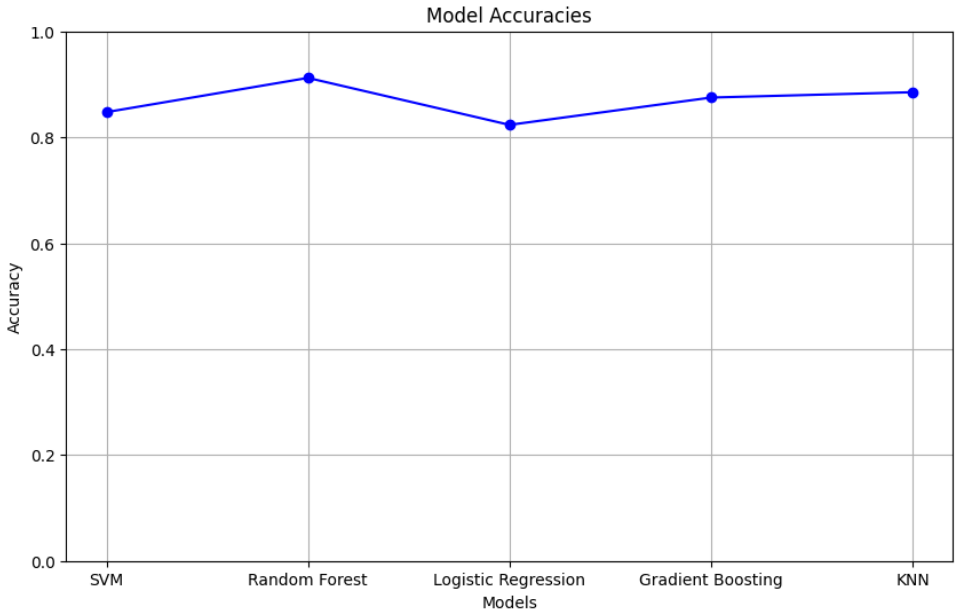


Figure 14: Model Accuracies

Figure 15: Model F1-score

Table 1: key findings

Models	**Accuracy**	**F1-score**	**AUC score**
Support Vector Machine (SVM)	84.74 %	0.8495	0.8480
Random Forest	91.21 %	0.9133	0.9127
Logistic Regression	82.34 %	0.8257	0.8239
Gradient Boosting	87.5 %	0.8766	0.8755
K-Nearest Neighbors Algorithm (KNN)	88.51 %	0.8912	0.8864

4 Results

After conducting experimentation we successfully developed a model that exhibits accuracy while utilizing only a minimal set of features. This accomplishment was made possible by employing the 'Select K Best' with the Chi-Square test function in combination with Random Forest for feature selection well as implementing the Random Forest algorithm for predictive modeling. The effectiveness of the Random Forest model proved to be exceptional as it provided importance scores that had an impact on our desired outcome. The outcomes clearly demonstrated that the Random Forest model outperformed all models we assessed. It achieved an accuracy rate of 91.21% an F1 score of 0.9133 and an AUC score of 0.9127. These outcomes suggest that the Random Forest model accurately predicted our desired outcome with a level of confidence solidifying its status as an efficient tool.

5 Future Work & Recommendations

We plan to incorporate additional data sources like ECG recordings or Medical images to further refine the model's accuracy. This model could be integrated into Electronic Health Records systems to provide real-time risk prediction at the point of care. By leveraging the power of Machine Learning, we can revolutionize cardiac arrest prediction and improve patient outcomes. We'll continue to invest in research and development in this.

References

[1] Cnn coronary heart disease prediction. 2020.

[2] An efficient convolutional neural network for coronary heart disease prediction. *Science Direct*, 2020.

[3] National health and nutrition examination survey. *Center for Disease Control and Prevention*, 2023.