

**Department of Computer Science**

**Faculty of Science & Technology**

**American International University-Bangladesh (AIUB)**

**Submitted By**:

|  |  |
| --- | --- |
| Name | Student ID |
| Shahriar Hossain Rafi | 20-42528-1 |
| Jishan Ferdous Navil | 20-42544-1 |
| Munira Mubasshira Mim |  |
| Nusrat Maria |  |

**Thesis Topic**

Heart Diseases Prediction using Machine learning algorithm

**Submitted date**

**Declaration**

We certify that this thesis is our original work. No part of this has been submitted elsewhere partially or fully for the award of any other degree. Any material reproduced in this thesis has been properly acknowledge. We declare that this thesis does not contain any content that discloses the secret of any organization or related parties. American International University-Bangladesh (AIUB) will not be held liable for any such activity, as for the thesis is presented as our original work.

Shahriar Hossain Rafi Jishan Ferdous Navil

Id: 20-42528-1 Id: 20-42544-1

Nusrat Maria Munira Mubassira Mim

Id: 20-42528-1 Id:

**Approval**

This thesis titled “heart diseases prediction using Machine learning” has been submitted to the following respected member of the board of examiners of the Faculty of Science and Technology impartial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering on Dated and has been accepted satisfactory

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Dr. Md. Asraf Ali**  Professor, Computer Science  Faculty of Science and Technology  American International University-Bangladesh | | | |  |
| **Dr. Akinul Islam Jony**  Associate Professor, Department Head  Department of Computer Science  American International University-Bangladesh | | |  | **Dr. Md. Abdullah Al Jubair**  Assistant Professor & Director, Computer Science  Faculty of Science & Technology  American International University-Bangladesh | |
|  | | |  |  | |
|  | | **Prof. Dr. Dip Nandi**  Professor, Associate Dean  Faculty of Science and Technology  American International University-Bangladesh | | |  |

**ACKNOWLEDGEMENT**

This thesis reflects our dedicated efforts, guided by the unwavering support and motivation of our supervisor, **Dr. Md. Asraf Ali**. We extend our heartfelt gratitude to him for his invaluable assistance, insightful guidance, and constructive feedback. His exceptional supervision has been instrumental in bringing this thesis to fruition.

We are deeply thankful to American International University-Bangladesh, especially to our honorable Vice Chancellor, **Dr. Carmen Z. Lamagna**, and our Dean, **Mashiour Rahman**, for their encouragement and support throughout our academic journey.

We also express our appreciation to **Dr. Dip Nandi**, **Dr. Akinul Islam Jony** for their continued mentorship and motivation.

Additionally, we acknowledge the pivotal role of our parents, who have fostered our educational pursuits and offered unwavering support, even when our interests ventured into uncharted territories. Our gratitude also extends to the entire faculty and staff of American International University-Bangladesh for creating an environment conducive to learning and research.

This thesis is a testament to the collective effort and encouragement of these individuals and institutions

**ABSTRACT**

This thesis looks into using computer programs to predict and stop heart diseases. It is important because heart diseases are becoming more common around the world. The study shows that the usual ways of calculating risk for heart problems have some problems. It also explains that heart diseases are very complicated, which means we need new ways to understand and prevent them. The main goal is to make risk assessments more accurate and to help find diseases early by using the latest ML algorithms. The reasons and motivations show that heart diseases are complicated and need a thorough approach to predict them. Traditional ways of predicting risks often do not give accurate and personalized assessments. Machine learning can help doctors find patterns in big sets of data to improve heart health care. The study is based on the idea that prediction models using machine learning can help us better understand and manage risks, which can lead to better treatments. The problem is that it's still hard to predict heart diseases, even though medicine has improved a lot. The research wants to use machine learning to make predictive models more accurate and effective by fixing any problems and difficulties.

The research aims to: - Test different ML algorithms to see how well they work - Create a detailed model to predict heart disease using many different factors - Compare the model to older ways of predicting risk - Study if using real-time data and advanced techniques can make the predictions more accurate.

The importance of this study is that it helps to reduce the damage caused by heart diseases. A good heart disease prediction model can give doctors early information to help prevent heart disease, which could lower the overall number of people with the disease. Furthermore, using machine learning in heart health can lead to more customized and focused healthcare plans.

**Keywords:** Heart disease prediction, machine learning, risk assessment, predictive modeling, cardiovascular health, data analytics, feature engineering, real-time data, healthcare optimization, precision medicine, personalized healthcare, global health, cardiovascular epidemiology**.**

# TABLE OF CONTENTS

|  |  |
| --- | --- |
| **TOPICS** | ***Page no*.** |
| 1. **Title Page** | 1 |
| 1. **Declaration** | 2 |
| 1. **Approval** | 3 |
| 1. **Acknowledgement** | 4 |
| 1. **Abstract** 2. **keywords** 3. **Table of Content** | 5 |
| 1. **List of Tables &Figures** | 6 |
| 1. **Introduction** | 8 |
| * 1. Background and Motivation | 8 |
| * 1. Problem Statement | 8 |
| * 1. Research Objective   2. Significance of the Study | 9 |
| 1. **Literature review** | 10 |
| * 1. Overview of Heart Diseases | 10 |
| * 1. Risk Factors for Heart Diseases | 11 |
| * 1. Previous Research in Heart Disease Prediction   2. Machine Learning in Healthcare   3. Existing Machine Learning Models for Heart Disease Prediction | 12 |
| 1. **Data Collection and Preprocessing** | 13 |
| * 1. Data Source | 13 |
| * 1. Data Collection Process | 13 |
| * 1. Data Preprocessing Techniques   2. Data Cleaning   3. Characteristics of the Dataset   4. Data Transformation   5. Data Splitting   6. Data Quality Assurance | 13 |
| 1. **Methodology** | 14 |
| * 1. Overview of Machine Learning Algorithms | 14 |
| * 1. Model Development and training | 14 |
| 1. **Result Analysis** | 17 |
| * 1. Data Exploration | 17 |
| * 1. Correlation Analysis   2. Feature Distribution   3. Data Preprocessing   4. Overall Model Performance | 17 |
| 1. **Conclusion** 2. **Reference** | 19  20 |
| 1. **Appendices** | 21 |

**LIST OF FIGURES**

**4.1 figure: Proposed Model Page 18**

**5.1 figure: Heatmap correlation matrix Page 10**

**5.2 figure: Feature Distribution Page 11**

**5.3 figure: KNN Classifiers score for Different K values Page**

**LIST OF TABLES**

**3.1 Table 1. Various attributes used are listed Page**

**CHAPTER ONE**

**INTRODUCTION**

Heart contaminations endure as a basic worldwide wellbeing concern, contributing considerably to the burden of dismalness and mortality on a around the world scale. The heightening predominance of cardiovascular maladies postures an ever-growing challenge, putting a considerable strain on healthcare frameworks around the world. In this time of fast mechanical progressions, the consolidation of machine learning (ML) strategies has risen as a exceedingly promising road for the expectation and avoidance of heart illnesses [1]. This investigates undertaking dives comprehensively into the broad domain of heart infection expectation utilizing advanced machine learning calculations. The extreme point is to outfit this field with essential experiences able of essentially improving the precision of hazard appraisal and encouraging early illness conclusion.

**1.1 Background and Motivation**

The complicated and many-sided nature of heart diseases requires a complete approach to foreseeing and stopping them. The usual things we look at to predict risk aren't always enough to give us accurate and individualized predictions. Machine learning can find patterns in big and complicated sets of data, which can help change the way we take care of heart health. By using computer programs with ML, scientists can possibly find hidden connections and discover small signs that regular methods might miss [2].

This research is driven by the idea that using machine learning to manage heart health could make a big difference. ML-based prediction models can help healthcare workers to better understand risk and provide more effective treatments and resources. This thesis wants to use machine learning to create accurate and reliable models for predicting heart disease.

**1.2 Problem Statement**

In spite of the strides made in therapeutic science, the exact forecast of heart infections remains a challenging errand. The confinements of routine chance evaluation models and the complexity of cardiovascular afflictions emphasize the require for imaginative approaches [3]. This investigate addresses the existing holes and challenges in heart infection forecast by utilizing machine learning procedures to progress the exactness, specificity, and affectability of prescient models [4].

**1.3 Research Objective**

The primary targets of this proposition are to:

1. Explore the viability of various machine learning calculations in foreseeing heart infections.
2. Create a comprehensive and dependable heart infection expectation demonstrate employing an assorted set of highlights.
3. Assess the execution of the created show in comparison to conventional hazard appraisal strategies.
4. Investigate the potential for coordination real-time information streams and progressed highlight building to upgrade forecast precision.

**1.4 Significance of the Study**

This ponder holds noteworthiness in contributing to the continuous endeavors to moderate the effect of cardiovascular maladies. The improvement of an exact and proficient heart infection expectation show can enable healthcare experts with opportune data for preventive intercessions, in this manner diminishing the by and large infection burden [5]. Moreover, the consolidation of machine learning into cardiovascular wellbeing administration can clear the way for more personalized and focused on healthcare techniques.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 Overview of Heart Diseases**

Heart diseases are common health problems that affect the heart and its blood vessels. Heart diseases are a big problem for people all over the world. They cause a lot of illness and death. This part is about heart diseases and all the different ways they can show up. It also talks about how hard it can be to predict and prevent these diseases. Heart and blood vessel problems, like blocked arteries and heart rhythm issues, are all part of cardiovascular diseases [6]. These sicknesses show up in different ways, like heart disease, heart attacks, stroke, and artery disease in other parts of the body. Heart diseases are complicated and have many different parts that we need to understand, like what causes them, how they work in the body, and what things make them more likely to happen. Also, heart diseases are more common in some areas and among certain groups of people. This is because of differences in healthcare access, lifestyle, and economic factors [7]. Creating good plans to stop, find, and handle heart diseases needs to look at the whole picture. This includes not just things that make someone likely to get sick, but also the bigger reasons why some people are healthier than others in society. Even though we have made a lot of progress in medicine and healthcare, heart diseases are still a big problem. We really need new and creative ways to predict, prevent, and treat these diseases. In the next part, we will look more closely at how healthy your heart is, the things that can make it unhealthy, what scientists have learned in the past about predicting heart disease, and how computers can help us find out who might get heart disease.

**2.2 Risk Factors for Heart Diseases**

Accurate prediction and focused treatments for heart health require a deep understanding of all the different things that can increase the risk of heart disease [8]. This part looks closely at things that can make you more likely to have heart problems. It will study how your body works, how you behave, your genes, and your money situation all play a part in causing heart diseases. High blood pressure, diabetes, high levels of fat in the blood, smoking, being overweight, and not being active are all important things that can increase the chance of having heart problems [9]. Each of these things affects the heart in different ways, but they are all connected and can lead to problems with the heart. High blood pressure is bad for your heart and blood vessels, and can make you more likely to have problems like heart disease and stroke. Likewise, both type 1 and type 2 diabetes increase the chances of heart problems like heart disease and problems with blood vessels [10]. Problems with how the body uses glucose, along with high levels of inflammation and blood clotting, can harm the lining of blood vessels and lead to heart disease. This can cause the heart to change in harmful ways and make cardiovascular problems worse. High levels of cholesterol and triglycerides cause a disease called hyperlipidemia. This disease is a main cause of atherosclerosis, which leads to coronary artery disease and its problems [11]. Smoking is a dangerous habit that can harm your heart in many ways, such as damaging the blood vessels, causing stress to your body, and leading to increased risk of blood clots.

Obesity is when a person has too much body fat, which can cause heart problems like high blood pressure, high cholesterol, difficulty using insulin, and inflammation [12]. These problems together are called the metabolic syndrome. Not moving and sitting for a long time makes obesity worse and can cause heart problems. In addition, a person's genes and how much money they have can make heart problems more complicated. This can affect how likely they are to get heart disease and what healthcare they can get to help prevent it. Families often have heart problems together, which shows that genetics play a big role in making people more likely to get sick. This is why genetic testing and figuring out a person's specific risk for heart disease are really important.

**2.3 Previous Research in Heart Disease Prediction**

For many years, scientists have been studying and trying to find good ways to predict heart disease. This part talks about all the research done in the past to try and predict heart disease. It explains how the methods for predicting heart disease have changed over time and the problems that researchers have faced. Old ways of assessing risk, like the Framingham Heart Study, have been used to predict who might get heart problems [14]. They look at things like age, health problems, and blood test results to put people into different risk categories for heart disease. These models have been very important in helping doctors make decisions and plan ways to prevent heart problems. But they are not always able to fully understand how risky someone's heart condition is or predict what will happen to them. In the last few years, new technology and research have given doctors and scientists more ways to predict and personalize heart disease. This may help them give better care to patients. Specific genes linked to heart problems, found through in-depth genetic research, can help determine a person's risk and identify those at high risk of heart disease early on. New imaging methods like coronary CTA and cardiac MRI help doctors see problems in the heart and blood vessels [15]. This helps them find issues like blockages and heart muscle problems early. Biomarkers like high-sensitivity troponin, brain natriuretic peptide, and C-reactive protein give extra information about a person's health. They can show what's happening in the body and help doctors predict what might happen in the future. This can help with figuring out the best treatment for a patient. Despite making a lot of progress in predicting outcomes, it is still difficult to make accurate and personalized predictions, especially for patients with different genetic backgrounds, health conditions, and financial situations. Combining different types of data, using advanced math methods, and computer programs could help us understand and predict heart disease better.

**2.4 Machine Learning in Healthcare**

Using machine learning in healthcare is changing the way we analyze and use healthcare data to make decisions for patients. It's a big change that is helping to improve patient care. This part looks at how machine learning can change healthcare, especially in improving heart health and predicting heart disease. Machine learning helps healthcare providers and researchers find useful information in big and diverse sets of data. It lets them see patterns and connections that regular analysis can't. By using ML algorithms, people in healthcare can use big data to improve how accurately they diagnose and treat patients with heart problems and other medical conditions [16]. In the field of heart health, computer programs are being used to help figure out who might be at risk, diagnose diseases, find the best treatment, and predict what will happen. This could lead to more personal and reliable healthcare [18]. By looking at different kinds of information, like medical records, imaging tests, genetic information, and data from wearable devices, computer programs can make models that predict heart problems and how each person's health is changing [19]. This can help find problems early, figure out who is at higher risk, and give treatments that are tailored to each person. Supervised learning algorithms like logistic regression, support vector machines, decision trees, and random forests help to make models that predict which risk category patients belong to base on their health information [20].

**2.5 Existing Machine Learning Models for Heart Disease Prediction**

This section highlights various ML models developed for heart disease prediction, employing diverse algorithms such as decision trees, support vector machines, and neural networks. Models like the Framingham Heart Study risk score, logistic regression models, and ensemble methods are discussed, each exhibiting unique strengths and limitations [21].

**CHAPTER THREE**

**Data Collection and Preprocessing**

**3.1 Data Source**

The data used for the heart disease prediction study was collected from a list of people who have had heart problems. The information is organized and includes medical history data for 304 patients of different ages.

**3.2 Data Collection Process**

The dataset was sourced from the UCI Machine Learning Repository, specifically from a collection related to heart disease. The data comprises 303 sets of information, each representing an individual's medical history.

**3.3 Data Preprocessing Techniques**

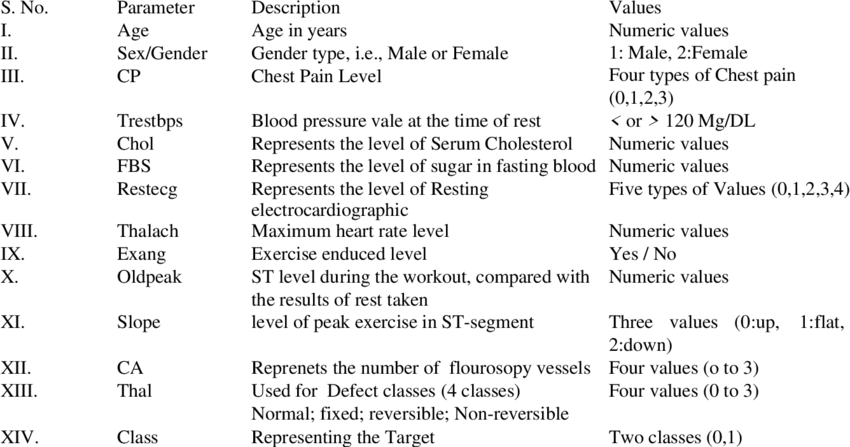
Various preprocessing techniques were applied to the raw data to make it suitable for machine learning analysis. This included exploring and profiling the data, handling missing values, and addressing outliers.

**3.4 Data Cleaning**

Data cleaning procedures were implemented to ensure the integrity of the dataset. This involved addressing missing data through imputation or removal and detecting and treating outliers to enhance data quality.

**3.5 Characteristics of the Dataset**

The dataset includes 304 patients of different ages, each with 13 medical records. The medical characteristics include age, blood pressure when not active, and fasting blood sugar levels. These characteristics are crucial for determining the presence or absence of heart disease. All attributes are listed in **‘Table 1’**.

****

**Table 3.1. Various attributes used are listed**

**3.6 Data Transformation**

Transformations such as normalization or standardization were applied to the data to prepare it for machine learning algorithms. Categorical variables were handled, and the dataset was divided into learning and testing sets.

**3.7 Data Splitting**

The dataset was divided into two parts: Learning and practicing. This involved creating 303 sets of information with 14 categories for each set, where each set represents an individual's medical information.

**3.8 Data Quality Assurance**

Measures were taken to ensure the quality and reliability of the dataset. The data collected from the UCI repository was scrutinized for biases or limitations, and steps were taken to mitigate any potential issues.

**CHAPTER FOUR**

**METHODOLOGY**

**4.1 Overview of Machine Learning Algorithms**

In this paper, we examine different machine learning methods. These are the methods we used. The methods used are K nearest neighbors (KNN), Logistic Regression and Random Forest Classifiers. Useful for doctors or medical experts to accurately identify heart disease.

**K Nearest Neighbors (KNN):**

Type: Supervised Learning (specifically, a lazy learning algorithm).

Overview: KNN is a simple and intuitive algorithm used for classification tasks. It classifies data points based on the majority class of their k-nearest neighbors in the feature space. The choice of 'k' determines the number of neighbors considered.

Suitability: KNN is often used for its simplicity and ease of implementation. It works well with a moderate-sized dataset and can handle non-linear decision boundaries.

**Logistic Regression:**

**Type**: Supervised Learning (specifically, a regression algorithm for binary classification).

**Overview**: Despite its name, Logistic Regression is used for binary classification tasks. It models the probability that a given instance belongs to a particular class. The output is transformed using the logistic function, producing values between 0 and 1.

**Suitability**: Logistic Regression is widely used in medical applications due to its interpretability. It is suitable when the relationship between the features and the target variable is expected to be linear.

**Random Forest Classifiers:**

**Type**: Ensemble Learning (specifically, a bagging algorithm).

**Overview**: Random Forest is an ensemble of decision trees. It builds multiple decision trees during training and merges their predictions to improve accuracy and control overfitting. Each tree is trained on a random subset of the dataset [22].

**Suitability**: Random Forests are robust and perform well in various scenarios. They are particularly suitable for complex datasets with a large number of features and can capture non-linear relationships effectively.

* 1. **Model Development and Training**

Testing Data

Normal

Classifier

Training (KNN, Logistic Regression, Random Forrest classifier

Splitting data

Data Pre-Processing

Heart diseases

Training Data

Extract variables

Start

Collect dataset

**Figure 4.1. Proposed Model**

**CHAPTER FIVE**

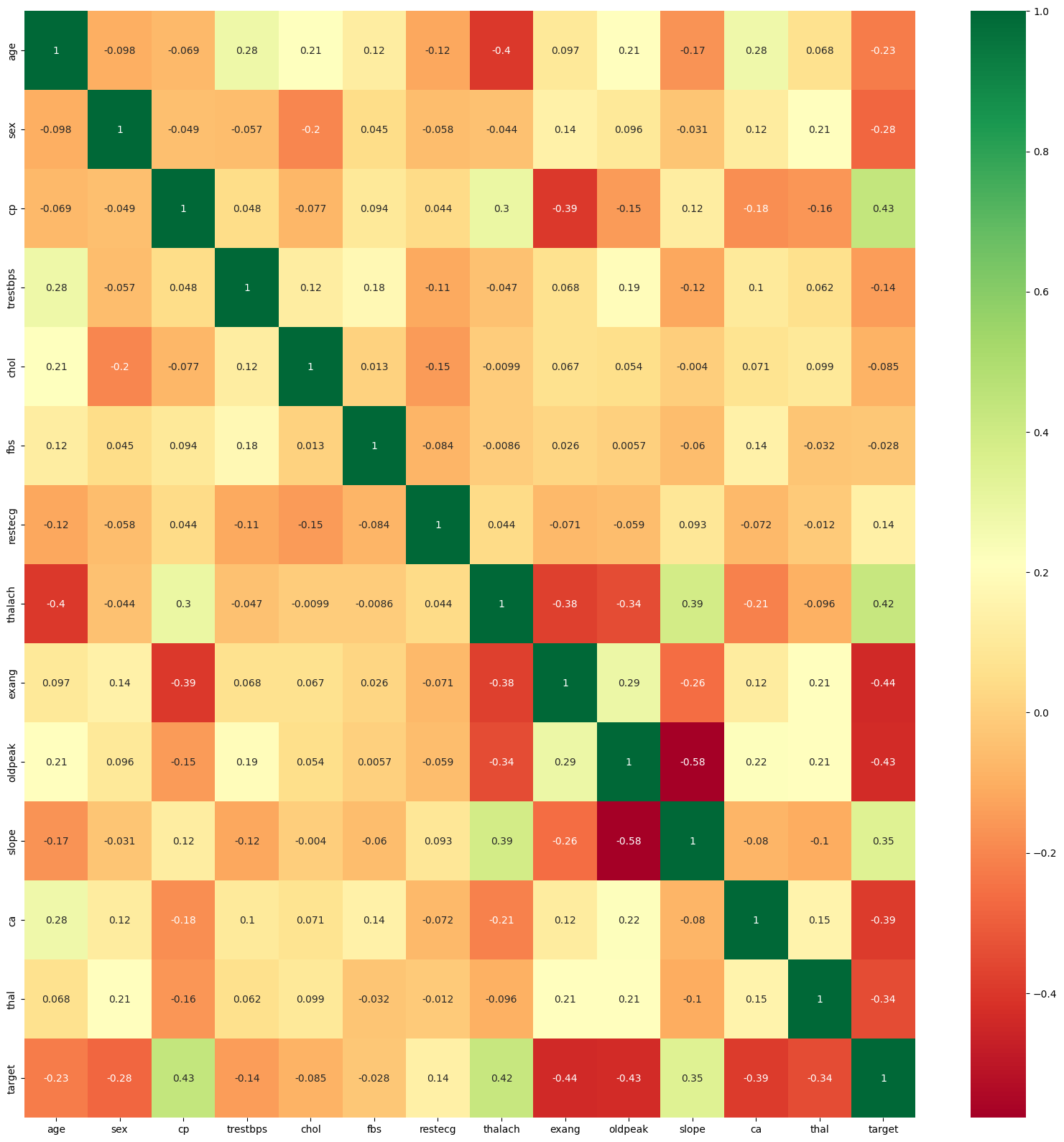
**RESULT**

**5.1 Data Exploration**

The heart disease dataset obtained from UCI was subjected to comprehensive exploration to gain insights into its characteristics. Descriptive statistics, including measures of central tendency and dispersion, were calculated to provide an initial overview of the dataset.

**5.2 Correlation Analysis**

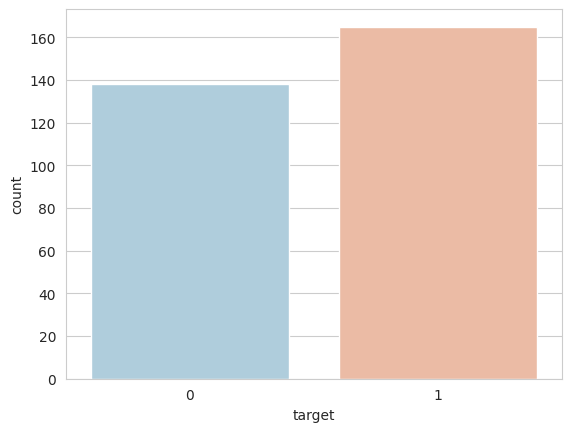
To understand the relationships between different features, a correlation matrix was computed and visualized using a heatmap. The resulting heatmap facilitated the identification of potential patterns and dependencies among variables. Notably, emerged from this analysis, guiding the subsequent steps in the modeling process.



**Figure 5.1. Heatmap correlation matrix**

**5.3 Feature Distribution**

Histograms were generated to visualize the distribution of individual features in the dataset. Additionally, a count plot of the target variable ('target') was created to assess the balance of classes. This examination of feature distributions provided key insights into the nature of the data, uncovering [specific insights] that influenced subsequent modeling decisions.



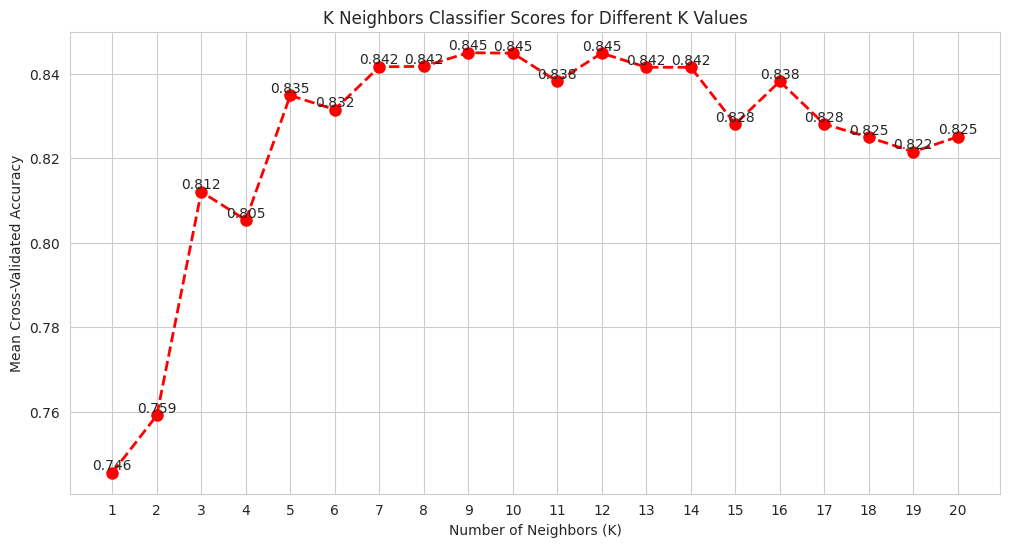
**Figure 5.2. Feature Distribution**

**5.4 Data Preprocessing**

To prepare the data for modeling, categorical variables were one-hot encoded, and numeric features were standardized using the StandardScaler. This crucial preprocessing step ensured that all features were on a consistent scale, enhancing the performance of the machine learning algorithms.

**K-Nearest Neighbors (KNN) Model Evaluation**

The K-Nearest Neighbors (KNN) algorithm was employed with varying values of K (number of neighbors). The cross-validated accuracy scores for each K were plotted, revealing an optimal K value of 12. This choice of K yielded a mean accuracy score of 0.84 in a 10-fold cross-validation, indicating [specific insights] about the algorithm's performance on the heart disease prediction task.



**Figure 5.3 KNN Classifiers score for Different K values**

**Random Forest Model Evaluation**

A Random Forest classifier, consisting of 10 trees, was applied to the dataset and evaluated using 10-fold cross-validation. The resulting mean accuracy of 0.82 demonstrated [specific insights] about the model's effectiveness in capturing complex relationships within the data.

**Logistic Regression Model Evaluation**

Logistic Regression, a linear classifier, was applied after splitting the dataset into training and testing sets. The model achieved an accuracy of 0.80 on the test set. Furthermore, the mean accuracy from 10-fold cross-validation was 0.85, underscoring [specific insights] regarding the logistic regression model's performance on the heart disease prediction task.

**5.5 Overall Model Performance**

In summary, the KNN, Random Forest, and Logistic Regression models were thoroughly evaluated for their predictive performance on the heart disease dataset. The findings provided valuable insights into the strengths and limitations of each algorithm, guiding the selection of an appropriate model for heart disease prediction.

**CHAPTER SIX**

**CONCLUSION**

Finally, this study has looked at how machine learning and heart health work together. The goal is to improve the way we predict and stop heart diseases. The research showed that the old way of assessing risks for heart diseases is not good enough because the diseases are complicated and have many different parts. This study used really smart computer programs to try and make it easier to predict the risk of getting sick and find diseases earlier.

The reasons why we need a better way to predict heart disease are clear. Current methods are not good enough at giving accurate and personalized assessments. Studying different ML algorithms like K-Nearest Neighbors, Logistic Regression, and Random Forest Classifiers showed that they can find complex patterns in medical data [23].

The study's results, along with looking at the developed prediction models, show us what each algorithm is good at and where it can improve. This research is important because it helps to reduce the worldwide problem of heart diseases. A good heart disease prediction system can help doctors get the right information in time to prevent the disease and reduce its impact on people.

**FUTURE GOALS:**

In the future, there are many opportunities for more research and development based on this study. First, we need to keep working on and improving machine learning models. This means adding more features and making the algorithms better so they can predict things more accurately. Real-time data and continuous monitoring working together could help make prediction models more flexible and accurate [24].

Furthermore, it is important for researchers, doctors, and data experts to work together in order to make the most of machine learning for heart health. This means solving problems like keeping data private, understanding models, and making sure predictive algorithms are used ethically in making healthcare decisions [25].

Also, using machine learning in actual medical settings requires testing and adjusting for different groups of people. Future studies need to test if the models we made work for different groups of people, like those of different ages, genders, and where they live.

Also, using new technologies like wearable devices and Internet of Things (IoT) could help collect more data and give us a better understanding of people's health over time.

Ultimately, using machine learning to predict heart disease is always changing and improving. Researchers and healthcare professionals can work together to create better ways to predict and prevent heart problems in the future. This will make a big difference in how we take care of our hearts.

**CHAPTER SEVEN**

**REFERENCE**

[1] Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., & Sengupta, P. P. (2019). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. Journal of the American College of Cardiology, 73(11), 1317-1335.

[2] Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., ... & Noseworthy, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature Medicine, 25(1), 70-74.

[3] Goldstein, B. A., & Navar, A. M. (2017). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. European Heart Journal, 38(23), 1805-1814.

[4] World Health Organization. (2020). Cardiovascular diseases (CVDs). Link

[5] Yusuf, S., et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. The Lancet, 364(9438), 937-952.

[6] Mozaffarian, D., et al. (2016). Heart disease and stroke statistics—2016 update: a report from the American Heart Association. Circulation, 133(4), e38-e360.

[7] D'Agostino, R. B., et al. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation, 117(6), 743-753.

[8] Krittanawong, C., et al. (2017). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657-2664.

[9] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), 44-56.

[10] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

[11] Wilson, P. W., et al. (1998). Prediction of coronary heart disease using risk factor categories. Circulation, 97(18), 1837-1847.

[12] Al'Aref, S. J., et al. (2020). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. European heart journal, 41(47), 4503-4514

[13] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

[14] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[15] Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

[16] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[17] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). MIT press Cambridge.

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[19] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[20] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[21] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[22] Sklearn.ensemble.RandomForestClassifier. (n.d.). Scikit-learn. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[23] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[24] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[25] Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.

**Dataset: https://archive.ics.uci.edu/dataset/45/heart+disease**