Mahfuzul Haque\textsuperscript{1}, Debasish Gupta\textsuperscript{2}, Aditi Golder\textsuperscript{3}, Jony Hossain\textsuperscript{4}, and Kishor Datta Gupta\textsuperscript{5}

\textsuperscript{1}Department of Computer Science and Engineering, University of Science and Technology
\textsuperscript{2}Department of Computer Science, Wake Forest University
\textsuperscript{3}Institute of Information Technology, Jahangirnagar University
\textsuperscript{4}Department of Cardiology, Zia Medical College Bogra
\textsuperscript{5}Department of Computer Science, Clark Atlanta University

January 22, 2024
CARDIA: Comprehensive Assessment Resource for Diagnosing, Interpreting, and Addressing Heart Diseases

Mahfuzul Haque¹, Debashis Gupta², Aditi Golder³, Dr. Jony Hossain⁴, and Kishor Datta Gupta⁵

¹Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology, Saidpur, Bangladesh
²Department of Computer Science, Wake Forest University, NC, USA
³Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh
⁴Department of Cardiology, Zia Medical College Bogra, Bangladesh
⁵Department of Computer Science, Clark Atlanta University, GA, USA

Abstract

Heart disease has been the leading cause of mortality globally. The necessity for quick access to trustworthy, dependable, and practical processes for early diagnosis and disease management pertains to numerous risk factors for heart disease. In the current global environment, detecting heart disease through early-onset manifestations is challenging. This has the potential to be fatal if not stopped in time. In isolated, semi-urban, or rural locations without access to heart specialists, accurate risk prediction and analysis may be essential for the early detection of cardiac issues. Artificial Intelligence (AI) and robotics are currently used in medical research. This addresses the urgent need for better ways to find, diagnose, and treat heart disease. To close the gap between theory and reality, we offer a dataset on cardiovascular disease that has been carefully put together. The variables in the dataset are age, gender, subtypes, symptoms, risk factors, and result variables that can be either 1 or 0.

¹Email: mahfuztajnim002@gmail.com
²Email: guptd23@wfu.edu
³Email: aditigolder2384@gmail.com
⁴Email: dr.jonyhossain@gmail.com
⁵Email: kgupta@cau.edu
information in the statistics comes from real hospitals and is clinical. AI models and robots can correctly diagnose heart problems using this vast data. They can also simultaneously handle a wide range of patient profiles and treatment scenarios. Our analysis shows that the dataset can be trusted and how thorough data collection and the use of medical books are linked. In this study, we proposed a new dataset named "CARDIA," which is based on real-life data collected from Bangladesh Hospital and shows how artificial intelligence and robots could change how heart disease is treated. This helps advance precise medicine and new ideas in the medical field.

Keywords — Dataset, Healthcare, Machine Learning, and Artificial Intelligence

1 Introduction

Heart disease, also known as cardiovascular disease, refers to a broad range of medical conditions that negatively impact the functioning of the heart\cite{4}. Heart diseases (HD), including heart failure, arrhythmia, and coronary artery disease, is still an indicative threat to world health and negatively affects people’s lives and healthcare systems significantly. According to recent data provided by the World Health Organisation (WHO), an estimated 20.5 million deaths occur annually on a global scale due to heart diseases and this figure accounts for approximately 32% of all recorded fatalities worldwide\cite{1}. According to a projection made by researchers, it is anticipated that by 2030, there will be an increase of approximately 24.2 million deaths annually\cite{7}. In addition, according to the European Cardiology Society (ESC), it has been reported that a significant number of individuals, approximately 3.6 million globally, are diagnosed with heart disease on an annual basis. This staggering figure translates to a cumulative total of 26 million cases\cite{5}. The allocation of healthcare resources for the treatment of heart diseases accounts for approximately 3% of the total expenditure. Furthermore, it has been observed that a significant proportion of patients diagnosed with heart disorders, precisely over 50%, experience mortality within a relatively short period of one to two years. This higher frequency of mortality is particularly notable in the Asian region\cite{10}. Due to their complexity, heart conditions are challenging to identify and treat. A few complex factors are essential to this task, including age, gender, symptoms, and risk features. Since there aren’t enough real-time datasets to accurately diagnose HD and treat it in a particular way, understanding these tiny signals is crucial for the early detection of cardiac abnormalities.

In this study, we presents a significant data set on cardiac illness that was extensively assembled with medical professionals working in hospital settings to satisfy this condition. Our dataset stands out from others since it used precise clinical data to diagnose any heart ailment, including symptoms, risk factors, age, and sex. It can be applied in real-world situations and is reliable and comprehensive. More specifically, The key contributions of this work include the following:
• The dataset consists of critical elements for accurately diagnosing cardiac disorders, including age and gender, which significantly impact disease progression.

• The strong correlations between symptoms and risk variables make our dataset unique. This enables classification algorithms to function effectively.

• A binary target variable is used to determine if the problem exists, as these components interact with different types of heart disease. The system utilizes comprehensive symptom and risk factor profiles to establish a person’s condition.

• References from a wide range of reputable clinical observations and medical research are incorporated into our collection.

• The dataset’s potential uses for diagnosing heart conditions on AI and robotics systems are carefully examined, showing the revolutionary potential for early detection and personalized treatments.

• A critical review of the ethical and practical effects of using AI and robotics in heart disease also shows how important it is to use data-driven methods to improve health care.

Our dataset on heart disease is an essential link between academic study and real-world applications in health care. By allowing AI models to spot and classify cardiac disorders, we see a future where the power of well-collected data will lead to faster diagnosis, more targeted treatments, and better patient outcomes.

The remaining paper is arranged as follows: Section II conducts a thorough literature review, Section III outlines the methodology in detail, Section IV describes the datasets, Section V rigorously analyzes the obtained results, and Section VI concludes the study and offers insights drawn from our findings.

2 Literature Review

Studies have demonstrated the prevalence of heart failure, arrhythmia, and coronary artery disease globally in recent years. Recent data show that heart disease is a significant cause of death and a global health risk. Most healthcare resources are spent on treating heart disease, underscoring its seriousness. Early detection and focused treatment are still challenging. This study fills a need by offering a comprehensive clinical sample. The distinct features of this dataset, like age-gender and symptom-risk factor associations, may change how AI and robots evaluate cardiac disorders. Consideration is given to the ethical and practical implications of AI and robotics in healthcare.
2.1 Comparison between datasets

This section explains what makes our dataset unique and how it differs from other datasets. Our collection is unique because it comes from real-world patient data. This means it is very similar to how cardiac detection and classification are done in the real world. In contrast to the datasets used so far, which include parameters like resting blood pressure, fasting blood pressure, resting electrocardiogram, slope, and old peak, our dataset consists of a wide range of essential characteristics. These things, like symptoms, risk factors, age, and sex, make classifying and finding different heart illnesses easier.

Traditionally, heart illnesses like ischemic heart disease, heart failure, valvular heart disease, and arrhythmias have been put into different groups. Ischemic heart disease often shows signs like angina, shortness of breath, and tiredness. Key risk factors include getting older, having high blood pressure, having high cholesterol levels, smoking, having diabetes, or having a family history of it. Heart failure causes tiredness, swelling, and trouble breathing. It can be caused by age, coronary heart disease, high blood pressure, and diabetes, among other things. Problems with the heart’s valves cause valvular heart conditions. They cause signs like murmurs, chest pain, and tiredness. Some of the causes are congenital disabilities and a history of rheumatic fever. Arrhythmia and irregular heartbeats can cause tremors, dizziness, and passing out. These changes can be caused by getting older, having a heart problem, or using drugs or alcohol.

Identifying these conditions requires a multifaceted approach that includes clinical assessments, medical records, physical evaluations, and various diagnostic tests like ECG, echocardiography, and serological analyses, all based on actual patient data. Our collection carefully records these essential details, making it a handy tool. The most important thing about the dataset is that it has dramatically affected the study of heart disease. This unique feature gets to the heart of our research, in which we offer an innovative dataset that makes a big step forward in analyzing heart disease.

2.2 ML and DL impacts on Heart Disease Classification

The topic of heart disease detection and prediction has been significantly impacted by machine learning (ML) and deep learning (DL) approaches. These models have been trained and evaluated using a variety of heart illness datasets, which has increased their accuracy in classifying various heart disorders. Using multiple datasets, the following impacts have been made, along with the related accuracy levels.

2.2.1 Cleveland Heart Disease Dataset

One of the most popular datasets in the field is the Cleveland Heart Disease dataset. Accuracy ranges from 70% to 85% when using ML algorithms like Decision Trees, Random Forests, and Support Vector Machines (SVMs). DL models with higher accuracies—between 85% and 90%—include Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs).
2.2.2 UCI Heart Disease Dataset

This dataset has also been heavily utilized. ML techniques with 70% to 80% accuracy include k-Nearest Neighbors (k-NN), Naive Bayes, and Gradient Boosting Machines (GBMs). Recurrent neural networks (RNNs) and extended short-term memory networks (LSTMs) are DL models that have further pushed the accuracy envelope, reaching 80% to 85%.

2.2.3 MIMIC-III dataset

A sizable critical care database has made it possible to make more reliable forecasts. Using ML algorithms and feature engineering, accuracy levels of 75% to 85% have been attained. With accuracy levels between 85% and 90%, DL models like Transformers and Gated Recurrent Units (GRUs) have increased the bar.

2.2.4 Framingham dataset

Dataset from the Framingham Heart Study: The Framingham Heart Study dataset has shed light on predicting cardiovascular risk. Accuracy levels for ML models using feature selection strategies ranged from 70% to 80%. DL models have shown better accuracy levels, ranging from 75% to 85%, by utilizing the temporal aspect of the data.

3 Method

3.1 Collection Process

The combined efforts of several Bangladeshi institutions compiled the data. The hospital personnel was briefed on the data collection process to ensure accuracy and consistency. It was decided to take the following measures:

3.1.1 Hospital Selection

To guarantee a geographically diversified patient group, hospitals from Bangladesh were selected. The selection of hospitals was based on their compliance with heart disease-related requirements. These included having ample facilities, a highly skilled medical team, and a sizable patient base.

3.1.2 Patient Enrollment

Patients diagnosed with heart conditions, who registered at the collaborative hospital at any point during the data collection period, were eligible for inclusion in the study. Participation was limited to individuals who had provided explicit consent for their data to be utilized in the research. It’s important to note that all data collection procedures strictly adhered to the guidelines outlined by
the Health Insurance Portability and Accountability Act (HIPAA) compliance regulations.

3.1.3 Data Variables and Collection Process
This data set has 19 symptom rows, each connected with one of four risk factors. Additionally, there are age, sex, disease, and target variables. The criteria were developed so that research could be conducted into the relationships between heart disease, symptoms, and risk factors.

Data collectors utilized conversations with patients, electronic health records, and medical records as the three primary sources of information to create patients’ profiles. The data collectors accurately recorded the patient’s demographic information, medical histories, symptoms, and risk factors.

3.1.4 Disease Confirmation
Heart diseases were diagnosed by cardiologists from the institution that was involved. Every diagnosis included in the dataset was made by trained professionals using accepted diagnostic criteria. The dataset used in this investigation is drawn from actual patient treatment records from Bangladeshi hospitals and extends through 2022. There are 329 data sets, each with complete patient records. The "target" variable, which shows if a patient has a heart illness, is the primary focus of the study. The analysis relies heavily on this binary variable. The "Age" column indicates the patient’s age, the "Sex" column indicates their gender, the "Disease" column shows the exact type of heart illness identified, and the "Target" variable indicates whether or not the patient has heart disease. Notably, four key risk factors in detecting the presence of specific heart illnesses are included in the dataset. More investigation and analysis of these interconnected risk factors should be included in the study. Hospital-based cardiologists validated the presence of heart disease in the dataset, proving that it accurately reflects the number of patients diagnosed and treated for the condition in Bangladesh in 2022. Knowing the signs and risk factors associated with various heart illnesses is crucial. The study paper will focus on these essential aspects and examine them in length. In addition, the dataset has 27 heart disease instances, which can be used to learn more about a wide range of cardiovascular disorders. Patients hospitalized for cardiac illness during the study’s four-month data collection period were included in the dataset to ensure accuracy.

3.2 Information Relation
This data set contains information about unique individuals and any prospective heart disease conditions that may have been present. The frequency of cardiac disease, as well as its demographics and age distribution, have been meticulously portrayed by the data. The aim variable classifies 75% of the samples as positive,
indicating the presence of heart disease, while the target variable classifies 25% of the samples as negative, indicating the absence of heart disease.

The presented material concludes with a comprehensive list of indicators and risk factors directly related to the categorization of heart disease. By examining these columns in conjunction with age and gender, it is possible to construct a predictive model that accurately diagnoses and categorizes heart disease. Considering the presence and severity of symptoms and risk factors, the target, which can be either one or zero, can be used to determine whether or not an individual in the dataset suffers from heart disease.

4 Dataset Description

The CARDIA dataset is publicly available on Kaggle, providing researchers with an opportunity to investigate a diverse dataset. The dataset can be accessed at [Dataset at Kaggle](https://www.kaggle.com/datasets/mahfuzulhaque22/heart-disease-detection-and-classification). The following subsections describe the features existing in our dataset.

4.1 Every Characteristic Description

To detect or classify heart disease using the provided 27-column dataset, it is necessary to comprehend the relationships between the columns and how they contribute to the diagnosis of heart disease. Let’s examine each column (symptoms, risk factors, age, and gender) to determine how they relate to heart disease classification.

4.1.1 Symptoms

The 19 symptom columns can reveal crucial information regarding the presence and severity of cardiac disease. Here is a breakdown of each symptom associated with cardiac disease:

The added data contains several columns that are required for this operation, and diagnosing or classifying heart disease is a critical medical duty. Because the data came from primary investigations, all of the symptoms and risk variables had binary values (1 = Yes, 0 = No). Let’s look at each symptom column individually and discuss how to diagnose or define cardiac disease and the relationships between them.

**Chest Pain:** Chest pain is a common sign of heart disease and can be caused by a number of conditions, including angina and coronary artery disease. It is brought on by constricted arteries that restrict blood flow to the cardiac muscle. Chest pain is essential for early detection because it identifies individuals at risk for heart-related events such as heart attacks.

**Shortness of Breath:** Having difficulty breathing is a typical symptom of heart failure, a condition in which the heart’s ability to circulate blood is impaired. It is crucial to evaluate this symptom in individuals with current
heart problems or a history of heart disease, as it is caused by fluid accumulation in the airways due to inadequate blood circulation.

**Heaviness or Tightness:** As with chest pain, chest heaviness or constriction may indicate angina or coronary artery disease. This sensation is caused by inadequate blood flow to the cardiac muscle, resulting in pressure or pain. It is closely associated with chest discomfort and is an additional sign of heart disease.

**Radiation into Arms, Neck, and Jaw:** Pain refers is agony that originates in the chest and spreads to the arms, neck, and jaw. It is frequently associated with cardiac issues. This symptom is a significant diagnostic indicator for heart-related disorders because similar neurological pathways cause it to be caused by similar neurological pathways.

**Congestion or Burning:** Swelling or burning sensations may result from fluid retention due to heart failure. These symptoms, which indicate poor cardiac function and fluid accumulation in multiple body regions, can aid in the diagnosis and treatment of heart failure.

**Abnormal Breathing:** Heart conditions can result in abnormal respiratory patterns, such as rapid or irregular respiration. These patterns may be the result of insufficient oxygen delivery to tissues or an irregular pulse, both of which are indicative of underlying cardiac abnormalities.

**Losing Flat Cause:** Understanding the significance of this word in diagnosing heart disease requires additional explication or context because it is not commonly used in mainstream medical terminology.

**Decrease of Strength:** Weakness or a loss of vigour may be an indicator of heart failure or a decline in cardiac function. The inability of the heart to circulate effectively can cause weakness by preventing vital organs and muscles from receiving sufficient oxygen and nutrients.

**Trouble with Balance:** Balance issues may result from a diminished blood supply to the brain due to cardiovascular disease or mini-strokes. The possibility that this symptom is related to specific cardiac rhythm disorders highlights the connection between heart health and neurological symptoms.

**Heart Rate Normal and Faster:** Deviations from the normal range of stationary heart rate necessitate additional investigation, as abnormal heart rates may indicate a variety of cardiac disorders. It emphasizes the significance of monitoring heart rate as a potential indicator of cardiac disease.

**Trouble with Swallowing:** Some forms of heart disease may cause difficulties with swallowing, especially if the esophagus is under stress from the enlarged heart. When assessing patients with known cardiac problems, this symptom may be significant.

**Low Blood Pressure:** Reduced cardiac output in heart failure and certain arrhythmias can both contribute to low blood pressure. It is a crucial factor to consider when assessing cardiac disease, as it can lead to syncope, dizziness, and poor organ perfusion.

**Missing in Heartbeat Rhythm/Abnormal Rhythm:** Heart disease is often associated with arrhythmias, or irregular heartbeats. These irregular
heartbeats may necessitate medical intervention and reduce the heart’s pumping efficiency.

**Need Pillow or Prefer to Sleep in Chair:** This symptom could be the result of uncomfortable heart problems when reclining flat. Patients with heart failure, for instance, may discover that sleeping on their backs makes breathing simpler.

**Syncopal Attack:** A syncopal attack occurs when a person unexpectedly loses consciousness or faints. It may be caused by a decrease in the brain’s blood supply, frequently resulting from structural or rhythmic cardiac issues.

**Debilitation:** In advanced cardiac disease, where the heart’s ability to function is severely compromised, debilitation or significant lethargy may occur. It is a warning sign for serious cardiac problems.

**Fever:** Fever is not always a distinct indication of heart disease, but it can indicate an underlying infection or inflammation that could harm the heart or its surrounding structures.

**Clubbing:** Expanding and curving one’s nails and digits is required for clubbing. Although it is a rare symptom of heart disease, it can develop in conditions characterized by persistently low blood oxygen levels or congenital cardiac abnormalities.

**Rash:** Heart disease and symptoms are not always related. Nonetheless, specific heart problems may occasionally generate cutaneous symptoms or responses as a side consequence.

These symptom columns are connected by their associations with heart disease and other symptoms. These symptoms, when combined with other risk factors such as age, gender, smoking, hypertension, and hypercholesterolemia, provide a complete picture for detecting and classifying heart disease. By comparing these columns, medical professionals can identify potential dangers, aid in early diagnosis, and implement appropriate treatment options for those at risk for or suffering from heart disease.

### 4.1.2 Risk Factor

The additional information in the four risk factors boxes can help determine risk factors can help find out if someone has heart disease.

**Smoking:** Smoking can cause damage to blood vessels and increase the risk of atherosclerosis, so it is well-known to be a factor that increases the likelihood of developing heart disease.

**Hypertension (High Blood Pressure):** One of the most significant risk factors for cardiovascular disease is high blood pressure because of its strain on the heart and blood vessels.

**Hypercholesterolemia (High Cholesterol):** When cholesterol levels are elevated, plaques are more likely to form in the arteries, raising the risk of heart disease.

**Myocardial Infarction (Heart Attack):** When there is a history of heart attacks in your family, there is a greater chance that you will experience cardiac issues.
4.1.3 Age

The patient’s age is one of the most important factors to consider when trying to get an accurate diagnosis. This is because the risk of developing cardiovascular disease rises with age. Age is critical in predicting risk and understanding disease evolution, and certain heart diseases are more common in those in their senior years.

4.1.4 Sex

Gender is essential to the condition because men and women may have different heart disease symptoms and risk factors. For instance, women may exhibit unusual symptoms that make it difficult to detect the disease early. Because of this, sex must be considered for a comprehensive review.

4.1.5 Target

In the objective column, a value of 1 indicates that heart disease is present, whereas a value of 0 indicates that heart disease is not present. To predict this variable, we will use the properties of the other columns.

4.1.6 Disease

Suppose a patient has already been diagnosed with heart disease. In that case, this column can tell you that information, which helps estimate the severity of the ailment and how long it will last.

4.2 Visualization

The visualization of some columns from our dataset is shown below. The primary columns are Target, Age, Sex, Disease, and Age. In the following breakdown, we have demonstrated how age, sex, disease, and target help identify the number of affected and unaffected individuals and those individuals’ ages and the relationship between age, sex, and illness in detecting heart disease. The prevalence of various diseases among persons of different ages.

The target variable shows whether or not a person has heart disease. It is a binary variable in which the existence of cardiac disease is indicated by a value of 1 and its absence by a value of 0. As previously noted, roughly 75% of the samples are given label 1 (indicating the existence of heart illness), and the remaining 25% are given label 0 (showing the absence of heart disease).

Age is a representation of the age of a person. The data indicates that those between 60 and 70 are the most affected group. People between the ages of 25 and 30 comprise the second-largest age cohort. Most affected individuals are in their 60s, those in their 50s, and finally, those in their 70s. In addition to the previously highlighted age distribution, the Heart Disease Dataset reveals several significant findings regarding age.
The dataset spans many life stages. It reveals age-related patterns and risk factors by showing heart disease prevalence across age groups. According to a study, age increases cardiovascular disease risk. Heart disease is most common in adults aged 60–70. This study confirms that heart disease risk increases with age. Even though the 60s have the highest percentage of affected people, a sizable chunk of the dataset includes younger people; heart disease can affect young people—the second-highest age group is 25–30. This shows the need for early detection, prevention, and lifestyle adjustments for more youthful heart health. Age is a substantial predictor of heart disease, according to the dataset. Understanding the relationship between age and cardiac disease may help predict models or statistical analysis. This dataset’s age distribution and heart disease correlation can help researchers and healthcare practitioners identify age-specific risk factors, focus interventions, and improve patient care.

The columns of the dataset have been used to account for a total of 28
Figure 3: Age

Figure 4: Disease based on Age
different diseases. For each of these disorders, the dependent variable has been coded with the value 1, indicating that it is present. The following provides some key details on how to identify and categorize different diseases.

This dataset contains a compendium of cardiovascular system symptoms and medical conditions. When lying supine, one condition, orthopnea, causes breathing difficulties. The dataset frequently documents "Fatigoa," or fatigue. Palpitations, or irregular or rapid heartbeats, are also observed. Aortic Root, Stenosis, and Regurgitation may indicate problems with the aorta. Angina and chest pain are two kinds of chest discomfort. A heart attack, or acute myocardial infarction, occurs when the heart loses blood. In addition, congestive cardiac failure is present. The heart valves are affected by rheumatic fever, resulting in rheumatic heart disease. Mitral regurgitation occurs when the mitral valve fails to close completely, allowing blood to flow back into the left atrium. Ulcers of the stomach or upper small intestine are peptic ulcers. Inflammatory pneumonitis is included in the dataset. Also included is pneumothorax or air between the lungs and thoracic wall. Hydrothorax is accounted for. "Previous Stroke," for instance, indicates a previous stroke, which occurs when the blood passage to the brain is obstructed. Anemia is described as decreasing hemoglobin and oxygen delivery. "Constriction," which may refer to blood artery constrictions that effect blood flow, is difficult to interpret without additional information. More information is required, but there may be an infection. People who are "Not Affected" have no symptoms. The dataset also includes "Minor Symptoms," which are unspecified faint or general symptoms. A thorough diagnosis requires...
This study analyses age, disease, and targeted results using a large dataset with many columns. The dataset requires age, gender, disease type, and goal variable. We analyzed and visualized the relationship between age, disease, and the goal. Our illustration shows how age and disease categories affect the final outcome. Age and disease are key to getting the desired outcome, and this article explains the many interrelationships in this multidimensional topic. The findings may enable more targeted and effective clinical treatment planning and decision-making.

5 Evaluation and Comparison

This section represents the F1-Score, Precision, and Recall comparison of the aforementioned models. Along with this, true-positive and false-positive rates have been shown here. The experiment has been conducted under cross-validation (CV) fold with 3, 5, 7, and 10. In the case of using SVM, we have tried its four kernel variants - i.e., linear, poly, radial basis function (aka rbf), and sigmoid.

5.1 Comparison of models

In Figure 8 BaggingClassifier, GradientBoostingClassifier, LogisticRegression, and RandomForestClassifier demonstrated more than 97% F1-score value out of them; the latter two classifiers showed superiority over the other two by maintaining the performance over several CVs. Out of the four kernel variation, linear, poly, and rbf performed almost the same, whereas sigmoid illustrated the lowest F1-score among all other classifiers.

Precision and recall ensure the reliability of the models shown in Figure 9. As Figure 9a describes, KNN and RandomForest Classifier achieved the high-
est precision followed by Logistic Regression, Gradient Boosting Classifier, and Bagging Classifier having an average 3% declination in percentage. In a variation of Kernels in SVM, rbf performs better in precision than in other settings. On the contrary, AdaBoost and Bagging Classifier demonstrated the highest 97% recall as shown in Figure 9b followed by the other classifiers. However, the marginal difference between the highest performers and others is minimal.

5.2 T-SNE and PCA comparison

T-SNE (t-distributed Stochastic Neighbor Embedding) and PCA (Principal Component Analysis) are both dimensionality reduction techniques used in machine learning and data visualization. They help in simplifying the representation of high-dimensional data while retaining important patterns and relationships between data points.

PCA is a linear dimensionality reduction technique that finds the principal components of the data, which are orthogonal (uncorrelated) directions that capture the maximum variance in the data. The first principal component accounts
for the most significant variance, followed by the second principal component, and so on. By projecting the data onto these principal components, PCA creates a lower-dimensional representation of the original data. t-SNE is a nonlinear dimensionality reduction technique that focuses on preserving local relationships between data points. It computes probabilities that represent pairwise similarities in the high-dimensional space and then tries to find a lower-dimensional embedding where these probabilities are preserved as much as possible. As a result, t-SNE is particularly useful for visualizing clusters or groups in complex datasets.

6 Conclusion

In conclusion, the healthcare sector is investigating robotics and artificial intelligence due to the urgent need for improved detection, diagnosis, and treatment of heart disease. Innovative strategies are required to improve patient outcomes and lessen the pressure on the healthcare system due to the destructive effects of cardiovascular disease on global health. To close the gap between theoretical study and real-world application, a carefully curated dataset on cardiovascular disease has been created. The data used in this study was collected in real hospital settings while being closely monitored by medical professionals. This dataset’s essential qualities, which include age, gender, heart disease subtypes, symptoms, risk factors, and a binary target variable, make it an invaluable resource for researchers and medical professionals. Its thoroughness makes it possible for AI robots and models to quickly and accurately diagnose and identify heart diseases, taking into account a wide range of therapeutic settings and patient characteristics. Combining precise data-gathering techniques with information from reliable medical literature is essential for validating the dataset’s
Figure 9: Precision Recall

Figure 10: Caption
reliability and future significance. This project’s ultimate goal is to improve patient outcomes and healthcare technology through the use of data-driven solutions. The establishment of the dataset, the data collection techniques, and the numerous potential this database gives for the study of robotics and artificial intelligence in the field of cardiovascular illness will all be covered in greater detail in the following sections of this research article. In order to diagnose and treat heart diseases, this research fills the gap between theoretical and practical applications of artificial intelligence (AI) and robotics. We have started this project to usher in a new era of precision medicine and healthcare innovation, significantly contribute to medical research advancement, and improve the quality of life for people with cardiac problems.

7 Acknowledgement

We would like to extend our sincerest gratitude to Dr. Jony Hossain for his invaluable contribution to the successful completion of this research. His tireless efforts in collecting and validating the dataset on heart-related conditions from Zia Medical College, Bogra were instrumental in the progression and accuracy of this study.

References


