Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

# MACHINE LEARNING-BASED INTELLIGENT APPROACH FOR PREDICTIVE ANALYTICS OF CARDIOVASCULAR DISEASE RISK ASSESSMENT

**[1]Hina Bashir,[2]Aneela Abdullah,[3]Bareerah Saeed,[4]Mubashar Hussain, [*5]Nasir Hussain**

*[1]Department of Mathematics and Statistics, University of Southern Punjab, Multan.*
*[2]College of Computing and Information Sciences (CoCIS), Karachi Institute of Economics and Technology (KIET), Karachi.*
*[3]Department of Computer Science, COMSATS University Islamabad, Vehari Campus*
*[4]Department of Computer Science , University of Engineering and Technology Lahore, Pakistan.*
*[*5]Department of Computer Science & IT, University of Southern Punjab Multan, Pakistan.*
*[*5]nasirhussain1192@gmail.com*

**Keywords**
Chronic Cardiac, Heart Attach, Machine Learning, and Classifier

**Article History**
*Received on 27 Feb, 2026*
*Accepted on 25 March, 2026*
*Published on 26 March, 2026*


**Corresponding Author:**
**Nasir Hussain**

**Abstract**
*Chronic cardiac diseases remain one of the leading causes of death across the globe, making early and accurate diagnosis critically important. Identifying these conditions in their early stages allows for timely treatment and can significantly reduce the risk of severe events such as heart attacks. This study focuses on using machine-learning techniques, applied to health examination data, to improve the prediction of chronic cardiac disease. Early detection of key risk factors—such as hypertension, smoking, high cholesterol levels, aging, male gender, and obesity—plays a vital role in preventing disease progression. In this follow-up research, statistical analysis is combined with machine learning methods to better understand and predict disease patterns. A clustering heatmap is used to visually group patients based on the severity of their condition, offering an intuitive way to distinguish between early-stage and advanced-stage cases. Patients in the early stages tend to cluster together, while those in more advanced stages are identified as high-risk and may require closer monitoring due to the potential for rapid deterioration in heart function. This integrated approach not only enhances predictive accuracy but also provides a practical tool for healthcare professionals. By offering clearer insights into disease progression, the proposed model can support physicians in making more informed and timely decisions when diagnosing and managing this complex and progressive condition.*

## 1. Introduction

The heart is the second largest organ in comparison to the brain, which is of higher priority in the Human body. Heart disease accounts for 16% of the world's total deaths. An estimated 30 to 40 percent of deaths in Pakistan are due to heart diseases. In living creatures, the heart plays a very significant role. (Marimuthu et al., 2018). The identification of the heat disease is problem due to unavailability of machine, limited medical staff and other valuable resources to verify the presence of this disease in developing countries. (Yahaya et al., 2020) By use of supervise learning, we can able to create agent base machine, which able to identify the behavior of certain problem independently. In supervised learning, the first step is to make learning about attributes of some problem with the specific output label. Following this, this train algorithm has been used for the identification of a class label of the test dataset.

The algorithm creates a mathematical model from a set of data that includes both the inputs and the desired outputs. Different machine learning approaches were applied to classify a new data set of cardiac disease. Different findings from experiments the best results from the SVM classifier when used the entire dataset (14 characteristics, 303 instances): only 48 examples were erroneously categorized. The naive Bayes and C4.5 algorithms produced comparable findings. However, they were inferior to the SVM. Based on the results, it can conclude that all classifiers performed admirably. On the data set, however, discovered that SVM outperformed both C4.5 and the naive Bayes classifier. (Chaki *et al.*, 2015). This system employs the Naive Bayesian and K-Nearest Neighbor algorithms, which are both data mining approaches. From the historical database, the system extracts patterns and relationships. This technique is helpful in hospitals for disease prediction. After testing, I discovered that the Naive Bayesian algorithm outperforms the KNN algorithm. To improve and expand the system, add other attributes. Other techniques, such as clustering, time series, and association rules used. Text mining can also use to mine data that isn't structured. Data mining and text mining can also combine. (B. and Priyadarshi, 2015)

There is a need for a practical feature selection method that identifies the significant traits contributing more to disease diagnosis. Particle Swarm Optimization (PSO) is one of the met heuristic algorithms used to discover the optimal solution in the shortest amount of time. The PSO algorithm identifies the more significant characteristics in a dataset and removes the irrelevant and redundant features. On the other hand, the standard PSO method has a problem picking the ideal weight to update the velocity and position of the particles. They also presented a new fitness function for PSO using the Support Vector Machine (SVM). (Vijayashree and Sultana, 2018). Data mining and machine learning are used to forecast the occurrence of cardiac disease. Analyze the performance of each algorithm's forecast and apply the system to the required area. Improve the accuracy of the algorithm by selecting more relevant features. Conclusion: From the literature review, they believe that creating a predictive model for heart disease patients is only marginally successful. That complicated models are required to boost the accuracy of early detection. The database gets more intelligent as more data is fed into it. (Marimuthu et al., 2018).

Data mining is used in hospitals to extract secret information that indicates the presence of some diseases. Various types of algorithms used that predict the heart diseases are playing crucial roles in the automatic detection of disease in hospitals. Support Vector Machine, Decision Tree, Nave Bayes, K-Nearest Neighbor, and Artificial Neural

Network are some machine methods used for cardiac disorders. (Chala Beyene, 2018).

A survey, begin by providing an overview of machine learning and providing brief explanations of the most often utilized classification approaches to diagnose heart disease. Then, in this subject, the review represents research papers on machine learning classification approaches. A complete tabular comparison of the surveyed papers is also offered. (Al-Janabi et al., 2018). The proposed model proves to be an efficient and effective one for accuracy improvement of the Naive Bayes classifier in which particle swarm optimization is implemented for feature subset selection, therefore providing comparable and even better performance in classification. They achieved that by developing an innovative algorithm to maximize the classification performance while decreasing features. The simulation results indicated that this method can automatically evolve a feature subset selection with fewer features and improve classification performance over the use of all of the elements in a dataset. (Dulhare, 2018).

This article covers data mining approaches in the prediction of heart disease. Heart disease is, by definition, a fatal disease. Several steps are taken to apply relevant methodologies in disease prediction. This report looked at research projects that used effective methods and were carried out by various researchers. Based on the comparison analysis, we can conclude that the SVM technology effectively predicts heart disease. It provides good accuracy by viewing various study articles. (Raju et al., 2018).

MAPO is a modified APO that tested on the heart disease real-world issue domain and predicts heart disease. Relational MAPO identifies optimal features from any relational dataset with great accuracy when the evaluation is on datasets like heart disease, MNIST, and Framingham. This method eliminates collinearity and dependency among related variables, which further enables the extraction of optimal non-collinear attributes from a relational dataset. The experiments indicate that video MAPO may be used for heart rate calculation from fingertip video. Hence, it may be used for accurate computation of heart rate in a person through fingertip video. (Sharma et al., 2020).

## 2. Literature Review

Clinical diagnosis highly depends on a doctor's skills. However, there are misdiagnosis and mistreatment cases. The patients are requested to participate in several tests of diagnosis. In most instances, not all the tests diagnose an illness. It is about forecasting heart disease using fewer parameters. Initially, it used 13 characteristics for heart disease prediction. A genetic algorithm discovered the most frequent heart problems, which reduced the number of tests needed by patients. 13 attributes were reduced to 6 Genetic searches. Subsequently, three classifiers are used to predict the diagnoses of patients with the same accuracy as above: Naive Bayes, Classification by Clustering, and Decision Tree. The decision tree outperforms the other two data mining algorithms combined with feature subset selection and model building time. After attribute reduction, Nave Bayes performs well as before. Clustering performs better than the other two methods. (M Anbarasi, E Anupriya, 2010).

The ensemble techniques used to boost classifiers using three powerful and leading methods such as bagging, boosting, and random subspace for the diagnosis of heart diseases. Compare and evaluate the performances of three prominent ensemble techniques designed for diagnosing valvular heart illnesses, thereby diagnosing valvular heart disease. In the study, a set of 215 samples were employed for validating the performance of the developed ensemble technique in the comparative research. The tests

indicate that ensemble classification methods are feasible and draw important conclusions about the efficacy of ensemble approaches in identifying valvular heart disease. (M Anbarasi, E Anupriya, 2010). An expert system was designed to identify heart illness using a (SVM) and the feedforward back propagation technique. This paper includes the detailed medical data along with preprocessing. The expert system used the (SVM) and feedforward Backpropagation techniques. 300 patients, 250 were employed to use as the training set and 50 use the evaluation process to make the system more authentic and dependable. Finally, they employed two neural network methodologies. However, the output was only 50% to 60%, indicating unreliable for the patient. Using other neural network approaches, this can use this expert system data to improve the accuracy of medicine. (Hannan et al., 2010).

Coronary artery calcium score improves risk classification in a typical risk factor-based prediction model. A total of 209 CHD events occurred during a median of 5.8 years of follow-up among a final cohort of 5878, with 122 of them being myocardial infarction, death from CHD, or resuscitated cardiac arrest. Model 2 compared with model 1 resulted in considerable improvement in risk prediction (net reclassification improvement=0.25; 95 percent confidence interval, 0.16-0.34; P.001). In model 1, 69 percent of the cohort was assigned to the highest or lowest risk categories, but in model 2, 77 % were highest or lowest risk categories. Using model 2, another 23% of those who experienced events was reclassified as high risk and another 13% who did not experience an incident was also reclassified as low risk. (Polonsky et al., 2010). An example of the neuro-fuzzy integrated system is shown below. From this study, it is clear that it is possible to identify risk based on the inputs from physicians. By using neuro-fuzzy integrated systems, in the process of getting an agreement with doctors' opinions, in NFIS training, the maximum classification accuracy in these systems has been achieved. Noisy voice recognition, noisy image filtering, nonlinear adaptive control, intelligent agents, and dynamical system performance analysis all gain advantages from the proposed research. The initial diagnosis of CHD by considering this proposed research activity through mobile SMS in remote locations, where doctors are not easily accessible. (Ansari and Gupta, 2011). The classification approach adopted a Multi-Layer Perceptron (MLP) using a Back Propagation learning algorithm along with a feature selection algorithm and biological test values. It used Thirteen characteristics to classify heart disease. Employ Information Gain to determine attributes, reducing the number of features collected from patients. Artificial neural networks are used to classify patient diagnoses. The number of qualities is reduced from thirteen to eight. For example, the percentage accuracy difference between 13 and 8 attributes in the training data set is 1.1 percent and in the validation data set is 0.82 percent. Khemphila and Boonjing, 2011). Rotation forest (RF) ensemble classifiers were built using 30 machine learning algorithms to examine their classification performance using Parkinson's, diabetes and heart disorders as illustrations from the literature. The datasets with three-dimension are first reduced using a correlation-based feature selection approach while testing. Second, the classification performance of 30 machine learning algorithms is computed for three datasets. Third, using the RF technique, 30 classifier ensembles are constructed to compare the performance of different classifiers using the same illness data. All of the tests are conducted using a leave-one-out cross-validation approach, along with the performance of 60 algorithms. In

the case of the diabetes, heart, and Parkinson's datasets, base classifiers led to 72.15 %, 77.52%, and 84.43% avg accuracy. (Ozcift ,Gulten, 2011).

## 3. Methodology

A prediction technique of cardiac diseases diagnosis is an important part of this chapter. To put a stop to Cardiac Disease has nowadays more than required. Quality of data-driven of health management for predict Cardiac Disease can uplift the entire research and safe process, so these health systems make sure that a large number of people can live good lives. So here Machine Learning (ML) comes into play. We will use the ML algorithms in the prediction of Cardiac Disease, and ML makes predictions quite accurate as well as easy. For feature selection, we have used the univariate selection and find out the best three features (Sweating, Vomiting, cholesterol) of our dataset. After that, we have used the Classifier property model (Extra Tree Classifier) for feature improvement. Then we have drawn a correlation matric heatmap to investigate our univariate selected feature with Extra Tree Classifier. In the subsequent step which is Data visualization, we have carried out some statistical analysis on our dataset like Gender V/s Target, Cholesterol V/s Target, Chest Pain V/s Target, ClusterMap, etc. We have performed the analysis of the Cardiac Disease patient dataset with proper data pre-processing. In pre-processing, we have checked the missing values, noisy data. Then, use different ML models were trained. The predictions are performed with Logical Regression, K-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine, etc. A detailed description of each step is given below. The proposed methodology has nine main steps and nineteenth sub-steps to predict Cardiac Disease. Figure 01 has shown the flow chart of our methodology.

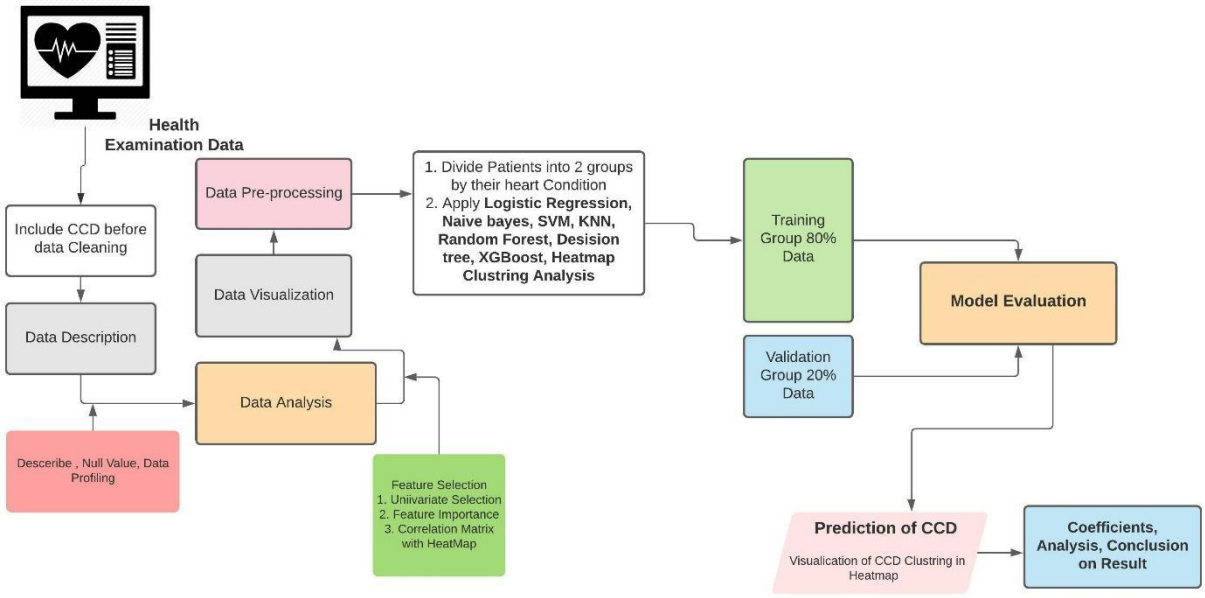


*Figure 1 Proposed Methodology*

### 3.1. Data Collection

We have collected the dataset from the city hospital; the name of the Dataset is **Heart Disease Data (Comprehensive)**. The heart disease dataset consists of **15** heart features (name, age, gender, chest pain, sweating, palpitation, dm, htn, smoking, family history, cholesterol, triglycerides, and one target class. Our dataset has a total of **379** samples of **Cardiac**

**Disease** patients**.** Below Figure 02 shows the details of a dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 378 entries, 0 to 377
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   name            378 non-null    object
 1   age             378 non-null    int64
 2   Gender          378 non-null    int64
 3   Chest Pain      378 non-null    int64
 4   Sweating        378 non-null    int64
 5   Palpitation     378 non-null    int64
 6   Vomiting        378 non-null    int64
 7   DM              378 non-null    int64
 8   HTn             378 non-null    int64
 9   smoking         378 non-null    int64
 10  Family history  378 non-null    int64
 11  Obesity         378 non-null    int64
 12  cholestrol      378 non-null    int64
 13  triglycerides   378 non-null    int64
 14  target          378 non-null    int64
dtypes: int64(14), object(1)
memory usage: 44.4+ KB
```

*Figure 2 Dataset Detail*

### 3.2. Dataset Description

In previous research, there have been many conflicts regarding the description of the metadata related to Cardiac Disease. As we know, there are various categories of metadata utilized in earlier studies. We have employed the two common metadata types shown below.

#### *3.2.1. Description – Meta Data - 01*

This type of Meta Data Description is a clean, well understand set of records. Anyhow the significance of some of the attributes is not much clear. Let's see the meaning of;

- Age: How much person's old now in terms of years
- Gender: Differentiate between Male and Female in term of (1 = Male, 0 = Female)
- cp: The experience of Chest Pain (typical angina is related to Value 01, atypical angina is related to Value 02, non-anginal pain is related to Value 03, asymptomatic is related to Value 04)
- chol: The Cholesterol amount in a person's body in terms of MG/DL
- restecg: This is the measurement of resting electrocardiographic in-person body (1 = normal, 2 = ST – T wave abnormality, 3 = probable)
- smoking: The smoking-related values are (Yes is 1, No is 0)
- target: This is the attribute, which classifies **Cardiac Disease** (Yes is 1, No is 0)

Our research use Description – **Meta Data 01** with the complete details of this attribute to understand it.

#### *3.2.2. Description – Meta Data -02*

- target: This is the identification of **Cardiac Disease**
  - Val 0: No
  - Val 1: Yes
- smoking: This is the smoking value
  - Val 0: No
  - Val 1: Yes

- cp: This is chest pain type
  - Val 1: Asymptomatic
  - Val 2: Atypical Angina
  - Val 3: Non-anginal Pain
  - Val 4: Typical Angina

| | age | Gender | Chest Pain | Sweating | Palpitation | Vomiting | DM | HTn | smoking | Family history | Obesity | cholestrol | triglyc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.000000 | 378.( |
| mean | 56.584656 | 0.605820 | 0.843915 | 0.328042 | 0.201058 | 0.193122 | 0.507937 | 0.515873 | 0.375661 | 0.246032 | 0.087302 | 167.431217 | 203.( |
| std | 13.180022 | 0.489322 | 0.363417 | 0.470123 | 0.401323 | 0.395271 | 0.500600 | 0.500410 | 0.484935 | 0.431268 | 0.282651 | 99.687479 | 150.8 |
| min | 24.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 23.000000 | 23.( |
| 25% | 48.000000 | 0.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 107.000000 | 78.( |
| 50% | 57.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 155.000000 | 147.( |
| 75% | 65.000000 | 1.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 1.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 228.750000 | 322.( |
| max | 91.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 681.000000 | 765.( |

***Figure 3 Describe Dataset Attributes***

**Figure 03** is showing the details of our Dataset in terms of the total number of records **(379)**, Find out the mean, Standard deviation, Min value, Max value**, 25%, 50%, 75%** Percentage of the Dataset.

*3.2.2.1. Null Value*

In this step, we need to identify the missing values in each attribute of the dataset. Figure 4 below illustrates that there are no null values present in any of the attributes.

```
name              0
age               0
Gender            0
Chest Pain        0
Sweating          0
Palpitation       0
Vomiting          0
DM                0
HTn               0
smoking           0
Family history    0
Obesity           0
cholestrol        0
triglycerides     0
target            0
dtype: int64
```

***Figure 4 Null values in Dataset***

We utilized a widely-used Heat Map tool to identify any null values in the dataset's attributes. To achieve this, we employed the Seaborn library to create a bar graph representing the null values. Figure 5 illustrates the results of the Heat Map, which indicates that there are no null values present in any of the dataset's attributes.

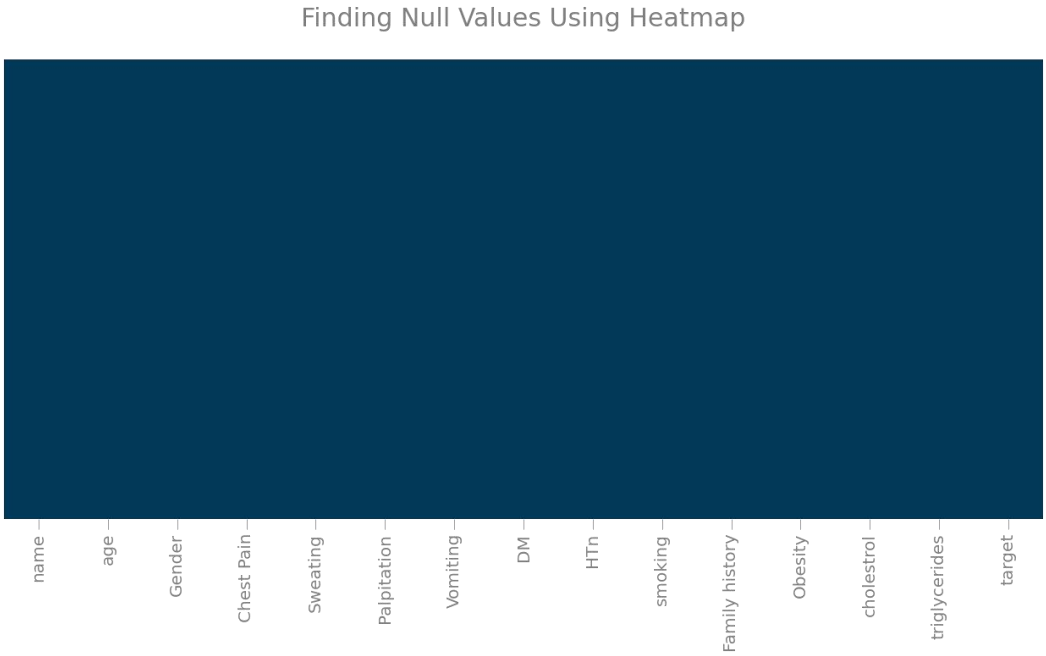


*Figure 5 Heat Map Result for Null Values*

#### *3.2.2.2. Data Profiling*

We have used the Pandas Profiling Library to further description of our dataset in terms of Overview, Variables, Missing Value, Correlations.

##### *3.2.2.2.1. Overview*

The dataset statistics indicate that there are no missing values in any of the attributes, resulting in a 0.0% missing value percentage. There are a total of 15 attributes, with 13 serving as input variables and 1 designated for classification or results. The ratio of duplicate rows stands at 5.5% of the total. Our dataset comprises two types of attributes: 14 with a numeric data type and 1 with a categorical data type.

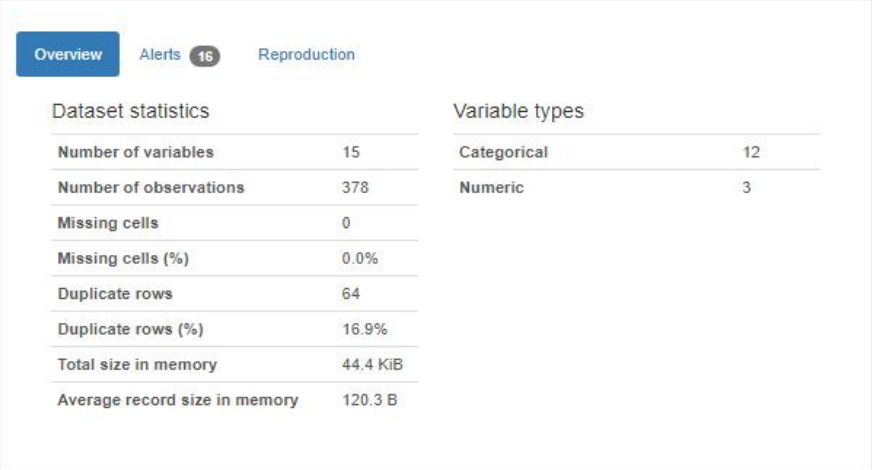


*Figure 6 Overview of Dataset*

### *3.2.3. Variables*

In this section, we have a detailed view of each attribute of the dataset.

##### *3.2.3.1.1. Age Attribute*

The data profiling indicates that the age attribute contains 56 unique records. The ratio of distinct ages is 13.5% of the total. There are no missing values in the age attribute, which stands at 0.0%. The average value for this variable is approximately 56.58465608.

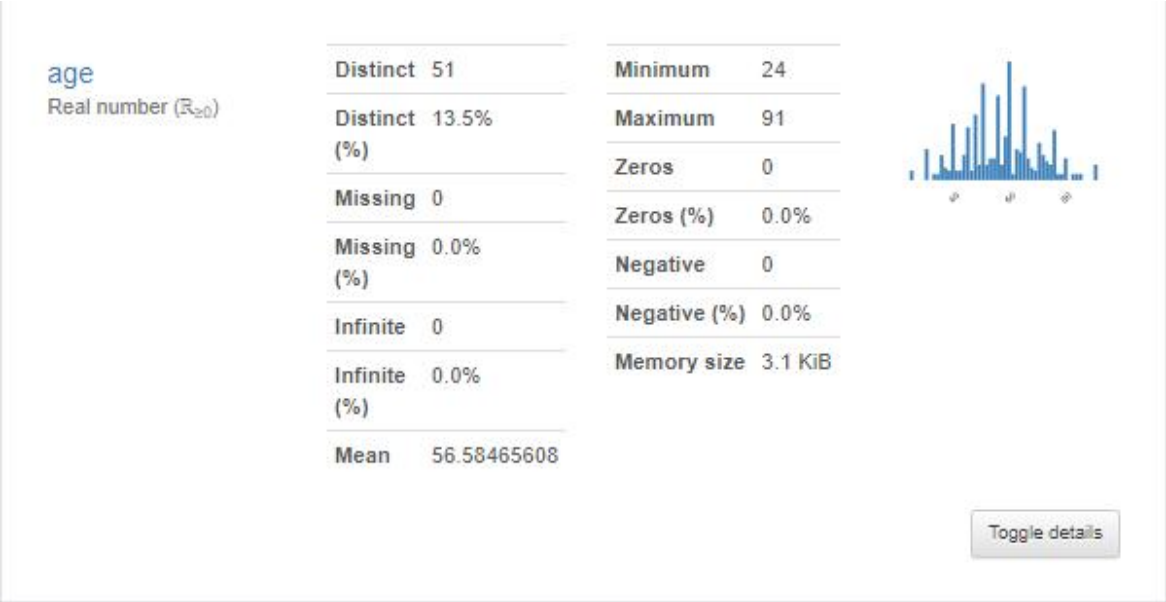


*Figure 7 Age Variable Details*

### *3.3. Sweating*

The data profiling indicates that the Sweating attribute contains 2 unique records. The ratio of distinct sweating is 0.5% to 100%. There are no missing values in the sweating attribute.


*Figure 8 Sweating Variable Details*

#### *3.3.1.1. Cholesterol*

The Data Profiling indicates that the cholesterol attribute contains 133 distinct records. The ratio of distinct cholesterol values is 35.2% to 100%. This cholesterol type is a numeric attribute, with values ranging from 681 to 23. There are no missing values in the cholesterol attribute, which stands at 0.0%. The mean value for this variable is 167.4312169.

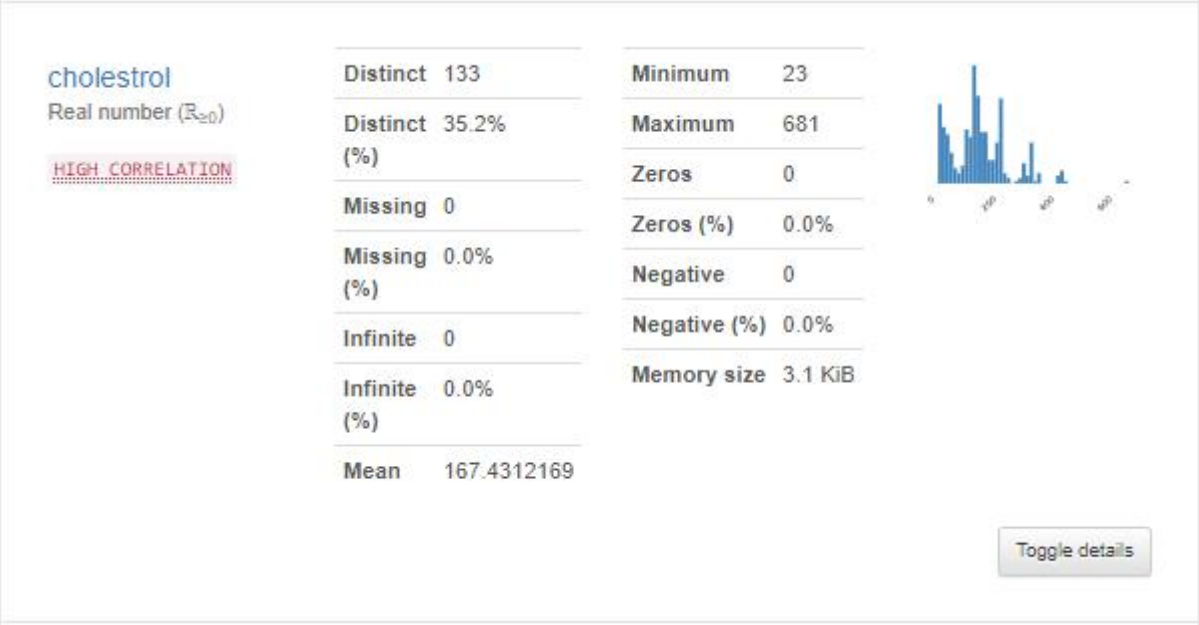


*Figure 9 cholesterol type variable Details*

### *3.3.2. Sex (Gender)*

The Data Profiling indicates that the Sex attribute contains 2 distinct records. The ratio of distinct Sex is 0.5% for females and 100% for males. This attribute is categorical, with 0 representing females and 1 representing males. Out of the 379 records, 229 are classified as type 0 (females) and 149 as type 1 (males).


*Figure 10 Gender variable Details*

#### *3.3.2.1. Palpitation*

The Data Profiling indicates that the Palpitation attribute contains 2 distinct records. The ratio of distinct Sex is 0.5% to 100%. Palpitation is a categorical attribute with two values: 0 and 1. The record 379 corresponds to type 1, while 76 records relate to type 0, totaling 302.

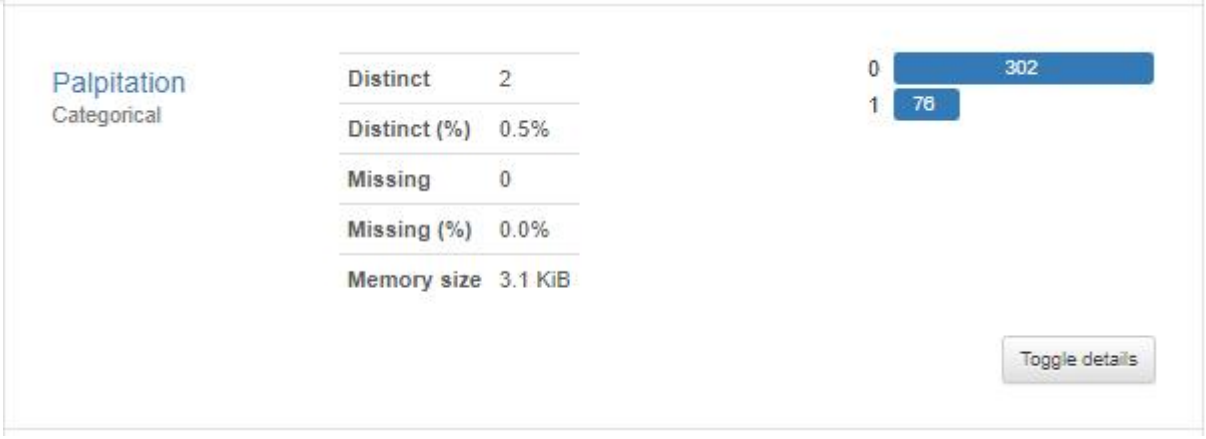


*Figure 11 Palpitation variable Details*

### *3.3.2.2. Vomiting*

The Data Profiling is showing that the vomiting attribute has a **02** distinct record in it. The ratio of distinct vomiting is **0.5%:100%**. The vomiting is a Categorical attribute. The **379** record is presenting type **1: 305** records is relating to type **0:73.**


*Figure 12 Vomiting Attribute Details*

### *3.3.2.3. Chest Pain (chest_pain)*

The Data Profiling is showing that the chest_pain attribute has a **02** distinct record in it. The ratio of distinct chest_pain_type is **0.5%:100%**. The chest_pain_type is a Categorical attribute. The **379** record is presenting type **1: 319** records is relating to type **0:59**.


*Figure 13 chest_pain Variable Details*

**a. Missing Values**

Mathematics and statistics define missing data as the absence of stored records in any attribute of a dataset during observation. Figure 15 illustrates that when 20% of the records are read from the dataset, which amounts to 75 out of 378, there are no missing values in any attribute. Similarly, when 40% of the records are read, totaling 151 out of 378, again, there are no missing values in any attribute. This pattern holds true for other percentages as well. Therefore, the data profiling of our dataset indicates that there are no missing data or values present.

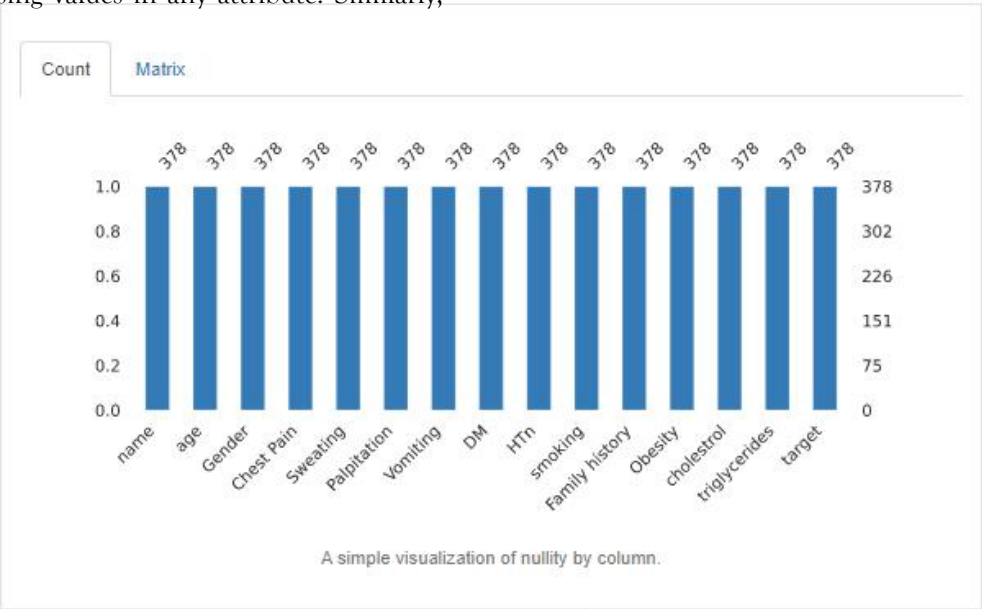


*Figure 15 Missing Values*

**b. Correlation**

We utilized the Phik (k) correlation coefficient to assess the relationship with the property of Cardiac Disease. Phik (k) serves as a tool for calculating correlation coefficients. In our analysis, we found that smoking has a nonlinear relationship with the target attribute. Conversely, factors such as chest pain, vomiting, perspiration, sex, age, palpitations, and obesity show a linear dependence in identifying Cardiac Disease. Phik (k) effectively measures the nonlinear dependencies among ordinal, numeric, interval, and categorical variables.

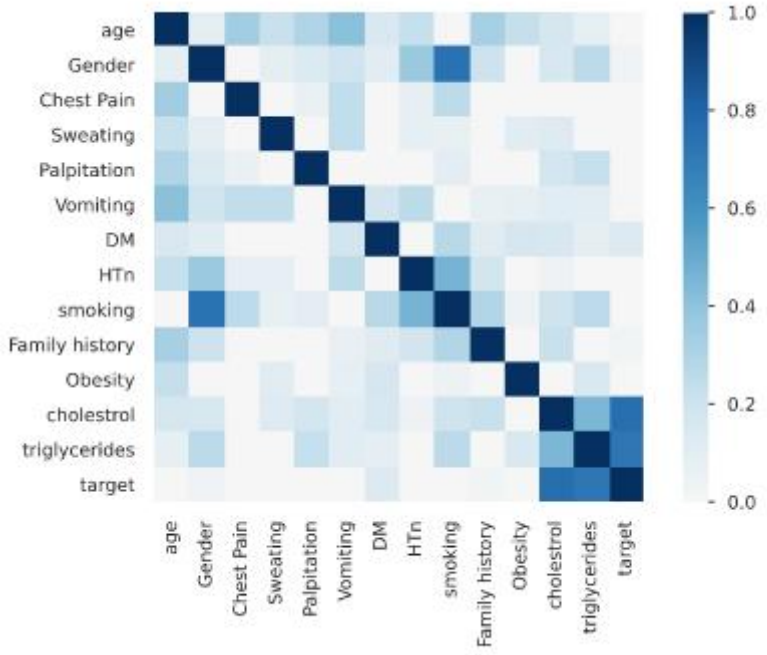


*Figure 16 Correlation between Data By using Heat Map Tool*

**c. Data Analysis**

This is our next step; we have employed a systematic approach and utilized statistical and conceptual techniques to illustrate, prove, describe, and condense our dataset, as well as to identify.

**d. Feature Selection**

To effectively build a successful model for Cardiac Disease, we need to apply machine learning techniques to reduce the number of input variables. This reduction is essential for lowering the computational cost and enhancing the model's performance. We can utilize methods such as univariate feature selection, feature enhancement, and a correlation matrix displayed as a heat map to illustrate our improvements.

**e. Univariate Selection**

This is a statistical test designed to identify key features that effectively correlate with performance variables. We utilized the SelectKBest class from the Scikit-learn library, which selects a specified number of the best features from a dataset. SelectKBest employs various statistical tests. Figure 21 illustrates a flowchart of our univariate selection methodology.

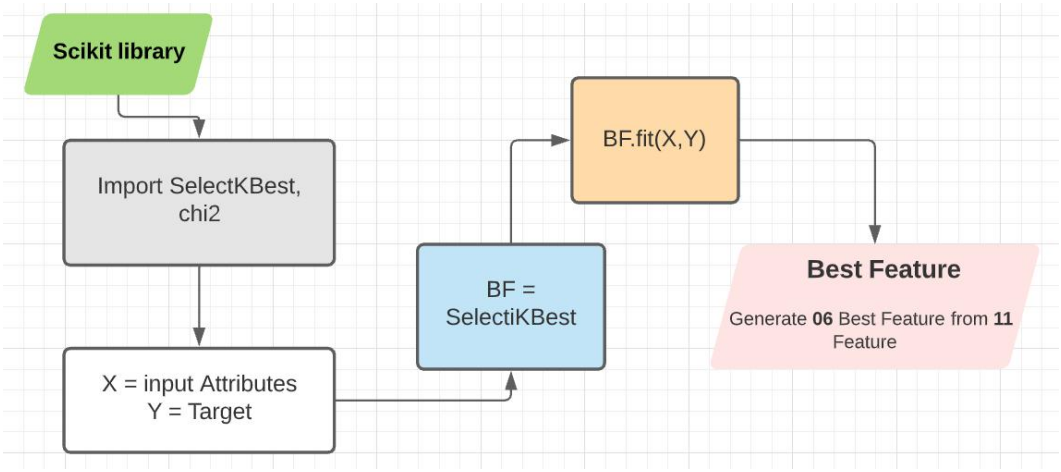


*Figure 17 Univariate Feature Selection Methodology*

After applying the above purpose univariate selection methodology, we have seen in Figure 16 that **DM, Obesity, Smoking, Family History, age, cholesterol** play an important role in reducing the model computational cost and also play role in improving model performance.

| | Specs | Score |
|---|---|---|
| 11 | cholestrol | 195.951406 |
| 6 | DM | 2.180041 |
| 9 | Family history | 1.139562 |
| 10 | Obesity | 1.136707 |
| 0 | age | 0.738296 |
| 8 | smoking | 0.631464 |

*Figure 18 Univariate Features*

**f. Feature Improvement**

To enhance our dataset, we utilized the model characteristics property. This allows us to significantly improve each feature within the dataset. The attribute value indicates the outcome for each attribute, meaning that a higher outcome signifies greater importance or usefulness in relation to the performance variable. We implemented the Extra Tree Classifier, which is a built-in class, to identify the top six features of our dataset.

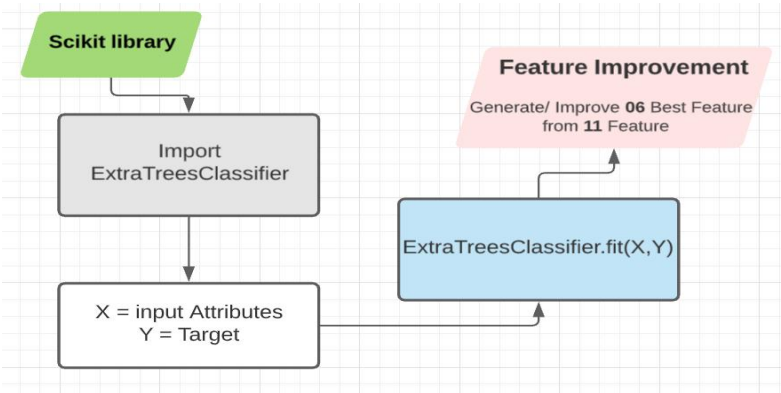


*Figure 19 Feature Improvement*

The below diagram shows that **cholesterol, age, HTn, Sweating, Palpitation, gender, DM, Vomiting** is improved as compared to early values.

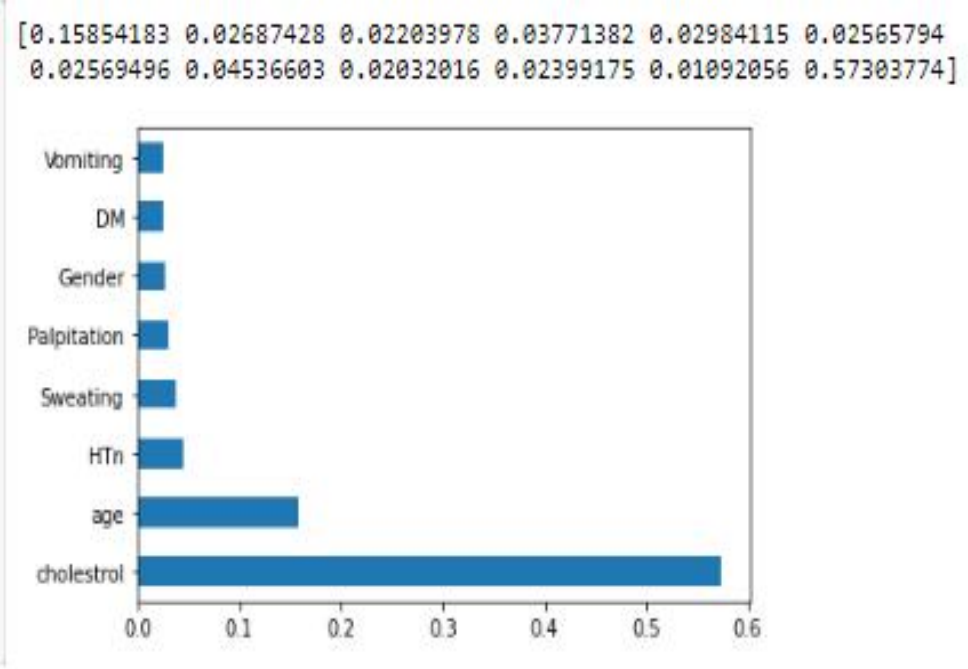


***Figure 20 Improve Features***

**g. Correlation Matrix with Heatmap**

Correlation shows how the attributes relate to the target attribute and to each other. A positive correlation means that as the value of the target variable goes up, so do the values of the features. Conversely, a negative correlation indicates that the target variable's value decreases when any of the feature values decrease. A heatmap effectively highlights which dataset attributes are most closely linked to the target attribute. We utilized the seaborn library to create the heatmap and visualize these relationships.

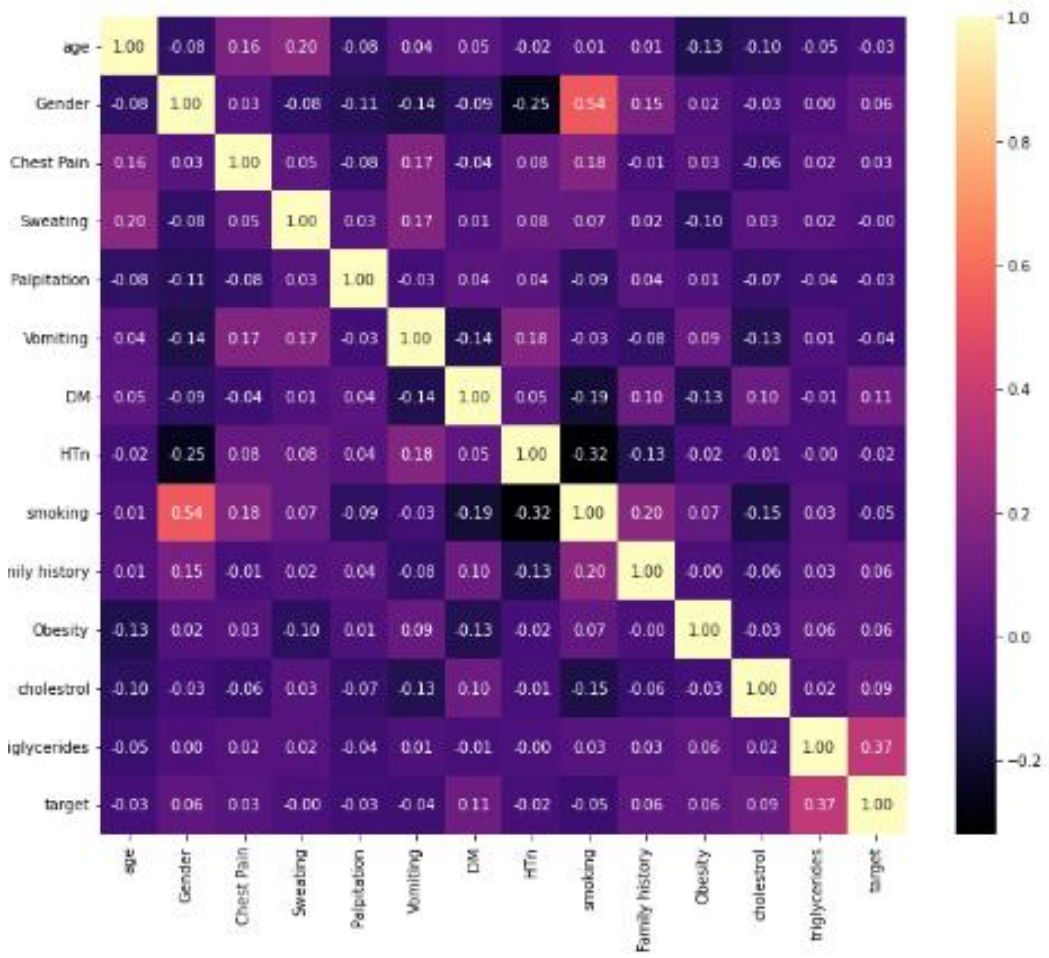


***Figure 21 Correlative Matrix with heatmap***

Figure 21 show that chest_pain_type, sex have a positive correlation toward the target attribute.

**h. Data Visualization**

In this step, we have illustrated various types of data visualization concerning the attributes of our dataset. We have depicted the relationships between cholesterol and resting blood pressure, chest pain and target outcomes, cholesterol and target outcomes, as well as a ClusterMap.

**i. Chest Pain Vs Target**

Figure 26 shows that most people with cardiac disease experience atypical angina chest pain. Another common type of chest pain found in patients is typical angina. Asymptomatic chest pain is a less common type of chest pain.

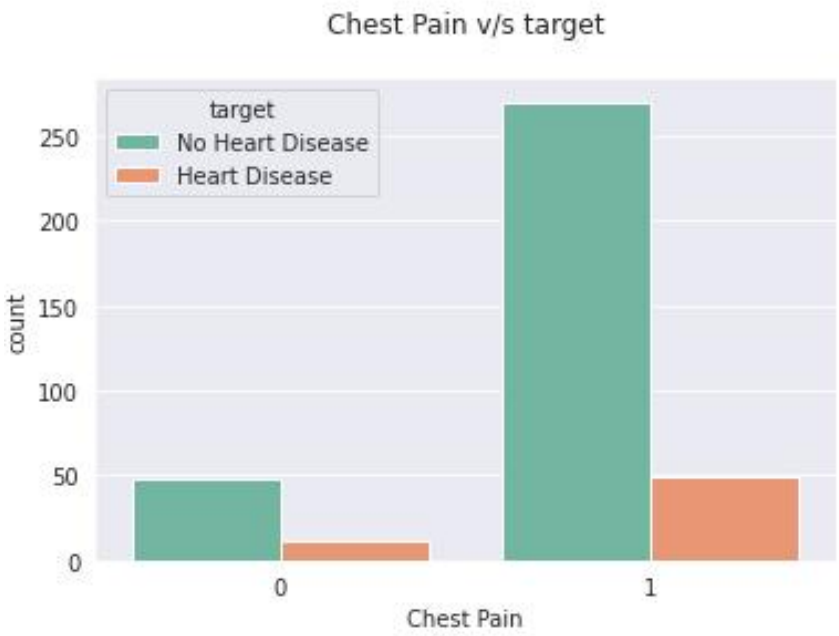


*Figure 22 Chest Pain Vs Target*

**j. Gender Vs Target**

Figure 23 shows that the high ratio of **Cardiac Disease** found in males as compared to females.

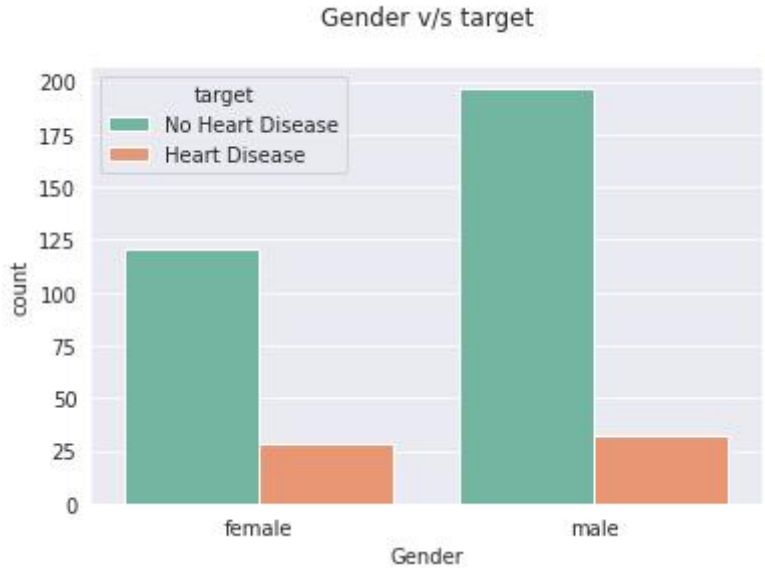


*Figure 23 Gender Vs Target*

**k. Sweating Vs Target**

Figure 24 shows that the high ratio of **Cardiac Disease** found in sweating.


*Figure 24 Sweating Vs Target*

**l. Cholesterol Vs Target**

Figure 25 shows that most of the people, who have **Cardiac Disease** suffering from **120** mg/dl to **140** mg/dl. People, who have cholesterol levels below **90** mg/dl are suffering from some


*Figure 25 Cholesterol Heart Disease Patients*

**m. Cholesterol Vs Age**

Medical scientists have evidence that when a person is dealing with multiple health issues, such as high cholesterol levels and age, these conditions can collectively heighten the risk of cardiac disease. Figure 26 below illustrates that if cholesterol levels fall between 120 mg/dl and 140 mg/dl, there is a significant likelihood that the individual may experience cardiac disease.

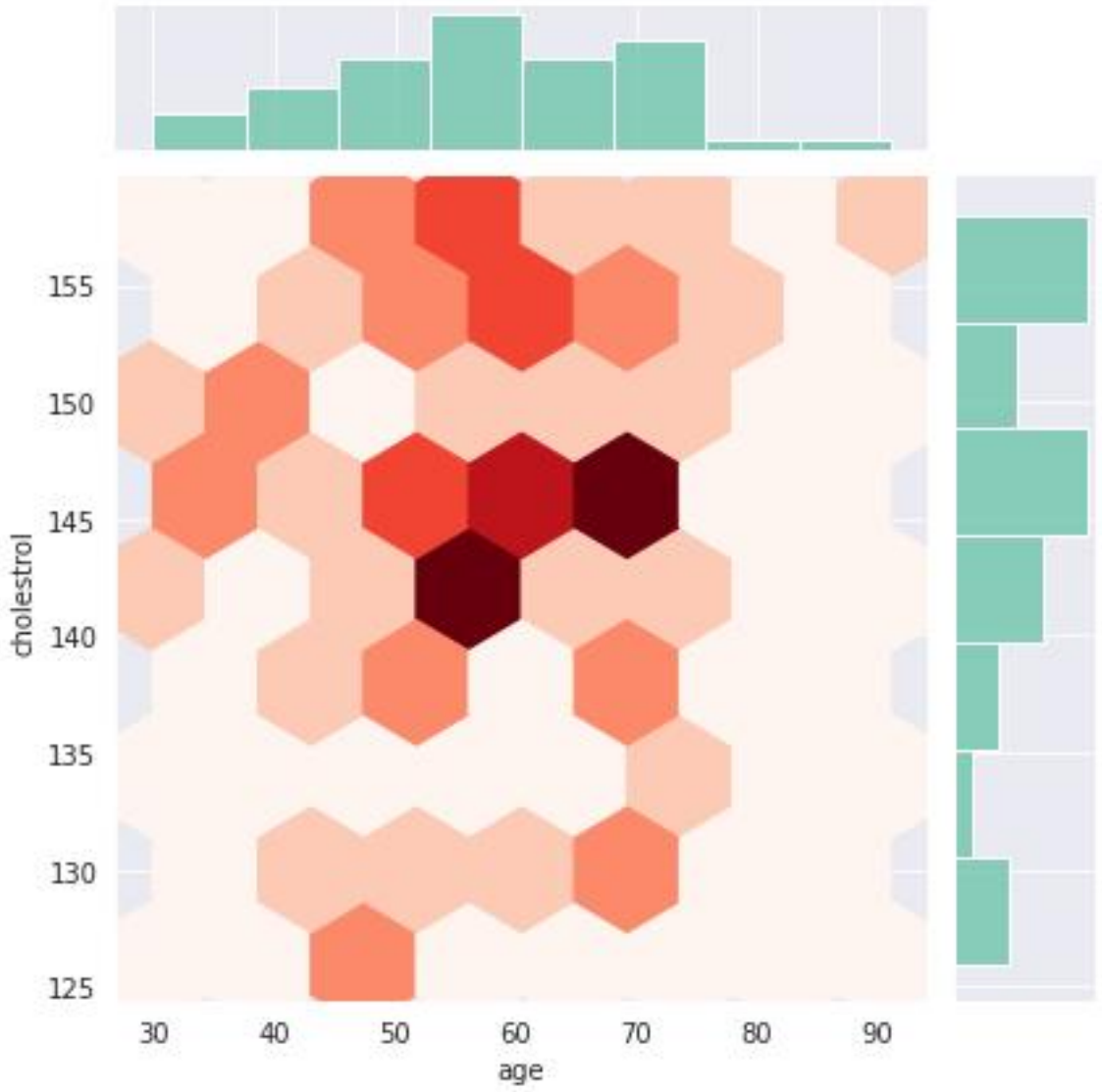


*Figure 26 Cholesterol Vs Age*

**n. Cluster Map**

The cluster map related to the target is illustrated in Figure 31. The cluster heatmap effectively highlights the dataset's attributes that are most closely linked to the target attribute. We utilized the seaborn library to create the heatmap showcasing these associated attributes. As shown in Figure 31, HTn, smoking, obesity, chest pain, and sex all exhibit a positive correlation with the target attribute.

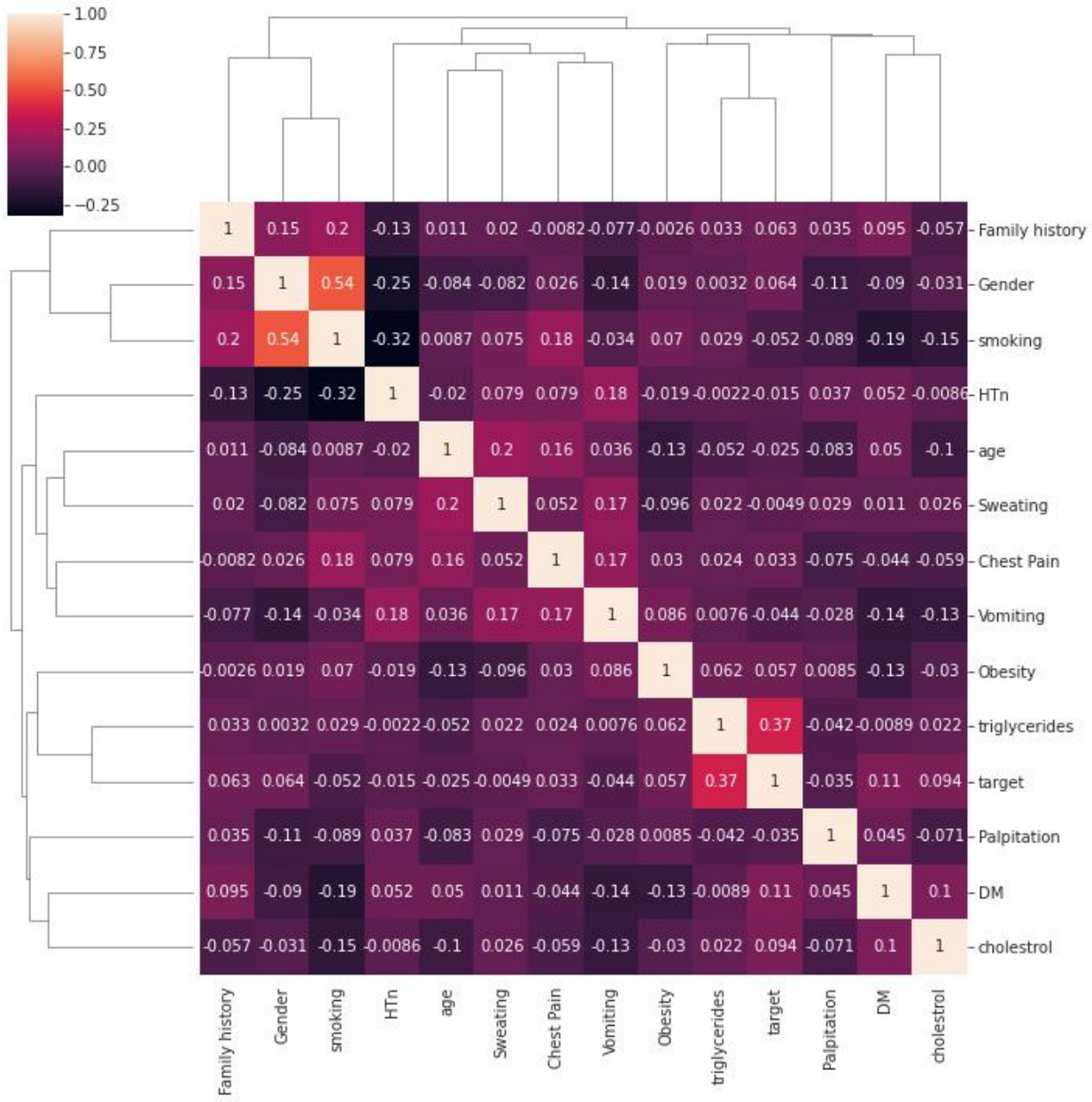


*Figure 27 Cluster Map*

A heatmap will be utilized to visualize the stages of patients identified through classification. This heatmap provides an interactive way to display individual values in a matrix, featuring color-coded grids and clustering for both columns and rows. Clustering will be employed to categorize a group of patients based on their health status. Patients will be sorted into various clusters using the hierarchical clustering method. (Wilkinson and Friendly, 2009)

**o. Pre-Processing:**

Preprocessing is a crucial step where we will exclude normal patients from the health examination data and focus on CCD patients using a specific algorithm. We will eliminate variables that have missing values and address any missing data using an algorithm if necessary. We will utilize various machine learning algorithms and compare their accuracy rates.

**p. Train & Test Data Split**

For data splitting, we have used the Scikit library and with the help of the train_test_split header file, we have divided the dataset into **80:20** ratios.

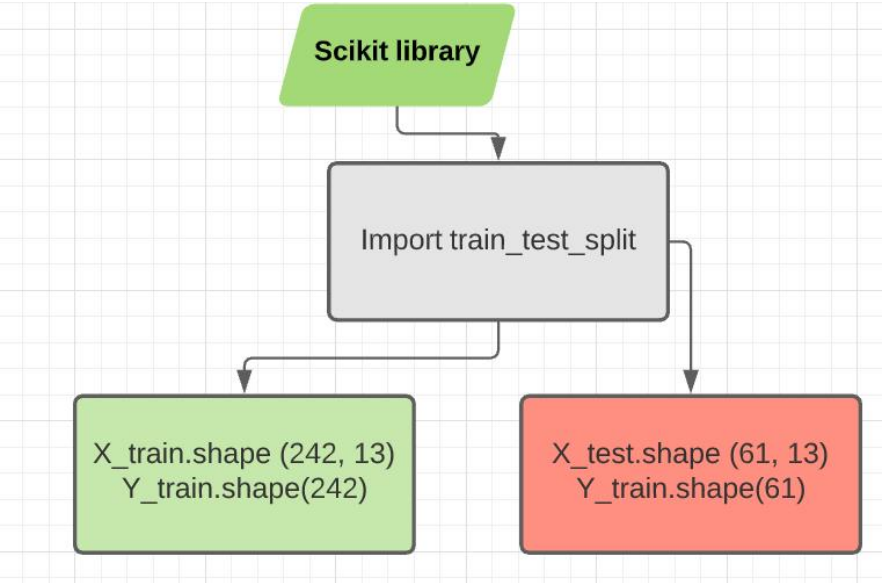


*Figure 28 Train & Test Data Split*

**q. Logistic Regression**

For Logistic Regression, we utilized the Scikit-learn library and, with the LogisticRegression module, we achieved an accuracy score of 81.58%.

**r. Naïve Bayes**

For Naïve Bayes, we utilized the Scikit library and, with the assistance of the GaussianNB header file, we achieved an accuracy score of 97.37% using Naïve Bayes.

**s. Support Vector Machine**

For our Support Vector Machine implementation, we utilized the Scikit-learn library and, with the SVM module, we achieved an accuracy score of 86.84%. In this experiment, we employed a linear kernel.

**t. K Nearest Neighbors**

For K Nearest Neighbors, we utilized the Scikit library and, with the assistance of the KneighborsClassifier header file, we achieved an accuracy score of 93.42% using Support Vector Machine.

**u. Decision Tree**

For our Decision Tree analysis, we utilized the Scikit library and employed the DecisionTreeClassifier. As a result, we achieved an impressive accuracy score of 97.37%.

**v. Random Forest**

For Random Forest, we utilized the Scikit library and, with the RandomForestClassifier module, we achieved an accuracy score of 100% using the Decision Tree method.

**w. XGBoost**

For XGBoost, we utilized the Scikit library and, with the assistance of the XGBClassifier header file, achieved an accuracy score of 96.05%. Figure 29 displays the graphical representation of the algorithms.

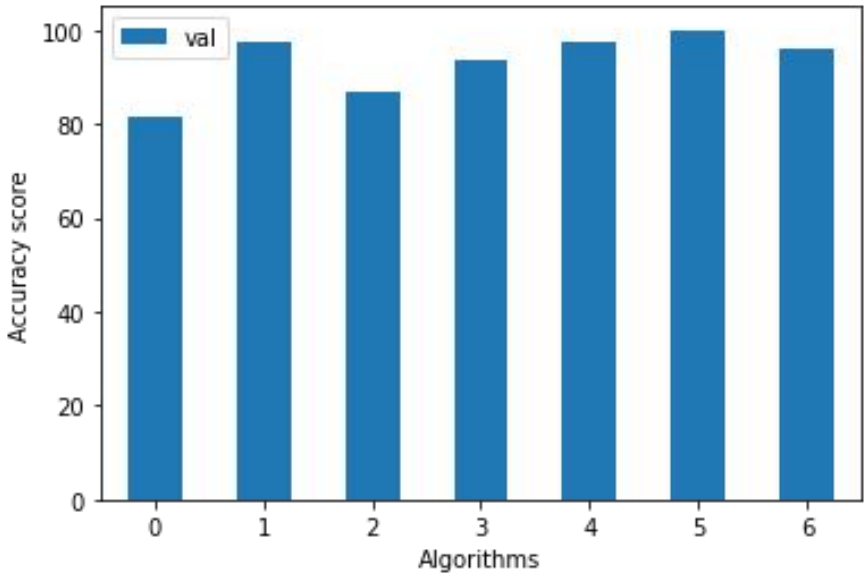


*Figure 29 Machine Learning Algorithm Comparison*

## 4. Discussion:

Our analysis highlights the effectiveness of predictive Machine Learning algorithms in developing health management systems and facilitating quick predictions of Cardiac Disease through component studies of a dataset. This study summarizes the results of various data visualizations, particularly focusing on cholesterol levels, maximum heart rate, chest pain types, resting blood pressure, and fasting blood sugar in relation to the target. We applied different Machine Learning algorithms to the dataset to identify the most effective predictive model. Our findings indicate that we need to assess the impact of specific publications that are statistically significant. The results illustrate the relationships between cholesterol levels and chest pain types, resting blood pressure and fasting blood sugar, among others, in the context of Cardiac Disease. The Random Forest algorithm has demonstrated superior performance in predicting Cardiac Disease compared to other Machine Learning methods. However, there is still room for improvement within the Random Forest algorithm. This issue is often referred to as publication bias, and it remains challenging to eliminate any research bias in today's landscape, as many factors contribute to the seriousness of this bias.

## 5. Conclusion and Future Work

We have utilized a range of statistical methodologies on our dataset, including Logistical Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, and XGBoost machine learning techniques. Among these, only the Random Forest model yielded satisfactory results. Our goal was to create an automated health care management model for predicting Cardiac Disease in patients, which we believe will effectively help in preventing such diseases. We performed various statistical analyses, such as Heatmap Clustering and Cholesterol vs. Target assessments, to pinpoint the most significant features associated with Cardiac Disease in patients. Our predictive model aims to support doctors and physicians in making accurate diagnoses of Cardiac Disease.

Looking ahead, we intend to explore which types of chest pain have a more pronounced effect on human health and what factors contribute to this. Besides the Cluster Heat Map model, we have also examined other models, though their effectiveness still needs to be confirmed.

Additionally, future research will focus on determining the appropriate levels of cholesterol, maximum heart rate, and Old Peak to mitigate the risk of Cardiac Disease.

**Bench Mark Comparison**

Table 03 presents a comparison of the proposed model's performance against earlier research in terms of Precision, Recall, F1-Measure, and Accuracy scores on the dataset. The findings reveal that our proposed model achieves an impressive accuracy ranging from 99.93% to 99.99%. In the study by Hannan et al. (2010), the authors utilized SVM, which resulted in 50% to 60% unreliable patient outcomes, thereby reducing the model's accuracy. K et al. (2018) employed various classification algorithms, including KNN, SVM, and ANN with modifications, leading to improved results compared to previous studies. Taylor et al. (2019) achieved an accuracy of 98.83% using Decision Tree, Random Forest, and Support Vector Machine algorithms. Mohan et al. (2019) introduced a hybrid HRFLM approach that merges the advantages of Random Forest and Linear

**Table 3:** ***Performance Metrics of Proposed Model***

| Data Set Name | Accuracy | Recall | F-Measure | Precision |
|---|---|---|---|---|
| Singh and Samagh, 2020 | 98.96 | 1.00 | 98.90 | 98.90 |
| Meng *et al.*, 2020 | 93.97 | 0.92 | 93.91 | 94.97 |
| Mohan *et al.*, 2019 | 94.96 | 0.93 | 96.50 | 96.42 |
| Proposed Model | 99.99 | 1.00 | 99.97 | 99.99 |

**References**

1. Khan, S. U. R., & Khan, Z. (2025). Detection of Abnormal Cardiac Rhythms Using Feature Fusion Technique with Heart Sound Spectrograms. Journal of Bionic Engineering, 1-20.
2. Waqas, Muhammad, Syed Umaid Ahmed, Muhammad Atif Tahir, Jia Wu, and Rizwan Qureshi. "Exploring multiple instance learning (MIL): A brief survey." Expert Systems with Applications 250 (2024): 123893.
3. Hekmat, A., et al., Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. Biomedical Signal Processing and Control, 2025. 109: p. 108040.
4. Khan, Saif Ur Rehman, Asif Raza, Inzamam Shahzad, and Ghazanfar Ali. "Enhancing concrete and pavement crack prediction through hierarchical feature integration with VGG16 and triple classifier ensemble." In 2024 Horizons of Information Technology and Engineering (HITE), pp. 1-6. IEEE, 2024.
5. Bilal, Omair, Arash Hekmat, Inzamam Shahzad, Asif Raza, and Saif Ur Rehman Khan. "Boosting Machine Learning Accuracy for Cardiac Disease Prediction: The Role of Advanced Feature Engineering and Model Optimization." The Review of Socionetwork Strategies (2025): 1-30.
6. Khan, M.A., Khan, S.U.R. & Lin, D. Shortening surgical time in high myopia treatment: a randomized controlled trial comparing non-OVD and OVD techniques in ICL implantation. BMC Ophthalmol 25, 303 (2025). https://doi.org/10.1186/s12886-025-04135-3
7. Ur Rehman Khan, Saif, Omair Bilal, Arash Hekmat, Inzamam Shahzad, and Asif Raza. "Advancing food safety: deep learning for accurate detection of bacterial contaminants." Memetic Computing 18, no. 1 (2026): 11.

8. Waqas, Muhammad, Zeshan Khan, Shaheer Anjum, and Muhammad Atif Tahir. "Lung-Wise Tuberculosis Analysis and Automatic CT Report Generation with Hybrid Feature and Ensemble Learning." In CLEF (Working notes), pp. 1-10. 2020.
9. Waqas, Muhammad, Muhammad Atif Tahir, Sumaya Al-Maadeed, Ahmed Bouridane, and Jia Wu. "Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (MIL) using vision transformer." Neural Computing and Applications 36, no. 12 (2024): 6659-6680.
10. Waqas, Muhammad, Muhammad Atif Tahir, and Salman A. Khan. "Robust bag classification approach for multi-instance learning via subspace fuzzy clustering." Expert Systems with Applications 214 (2023): 119113.
11. Khan, S.U.R., Asif, S., Bilal, O. et al. Lead-cnn: lightweight enhanced dimension reduction convolutional neural network for brain tumor classification. Int. J. Mach. Learn. & Cyber. (2025). https://doi.org/10.1007/s13042-025-02637-6.
12. Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Xiao, C. (2026). ShallowMRI: A novel lightweight CNN with novel attention mechanism for Multi brain tumor classification in MRI images. Biomedical Signal Processing and Control, 111, 108425.
13. Waqas, Muhammad, Muhammad Atif Tahir, and Rizwan Qureshi. "Deep Gaussian mixture model based instance relevance estimation for multiple instance learning applications." Applied intelligence 53, no. 9 (2023): 10310-10325.
14. Waqas, Muhammd, Muhammad Atif Tahir, and Rizwan Qureshi. "Ensemble-based instance relevance estimation in multiple-instance learning." In 2021 9th European workshop on visual information processing (EUVIP), pp. 1-6. IEEE, 2021.
15. Khan, M. A., Khan, S. U. R., Rehman, H. U., Aladhadh, S., & Lin, D. (2025). Robust InceptionV3 with Novel EYENET Weights for Di-EYENET Ocular Surface Imaging Dataset: Integrating Chain Foraging and Cyclone Aging Techniques. International Journal of Computational Intelligence Systems, 18(1), 1-26.
16. Khan, S.U.R., Zhao, M. & Li, Y. Detection of MRI brain tumor using residual skip block based modified MobileNet model. Cluster Comput 28, 248 (2025). https://doi.org/10.1007/s10586-024-04940-3
17. Al-Khasawneh, Mahmoud Ahmad, Asif Raza, Saif Ur Rehman Khan, and Zia Khan. "Stock Market Trend Prediction Using Deep Learning Approach." Computational Economics (2024): 1-32
18. Khan, U. S., & Khan, S. U. R. (2025). Ethics by Design: A Lifecycle Framework for Trustworthy AI in Medical Imaging From Transparent Data Governance to Clinically Validated Deployment. arXiv preprint arXiv:2507.04249.
19. Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Xiao, C. (2026). ShallowMRI: A novel lightweight CNN with novel attention mechanism for Multi brain tumor classification in MRI images. Biomedical Signal Processing and Control, 111, 108425.
20. Khan, Z., Khan, S. U. R., Bilal, O., Raza, A., & Ali, G. (2025, February). Optimizing Cervical Lesion Detection Using Deep Learning with Particle Swarm Optimization. In 2025 6th International Conference on Advancements in Computational Sciences (ICACS) (pp. 1-7). IEEE.
21. Waqas M, Bandyopadhyay R, Showkatian E, Muneer A, Zafar A, Alvarez FR, Marin MC,

Li W, Jaffray D, Haymaker C, Heymach J. The Next Layer: Augmenting Foundation Models with Structure-Preserving and Attention-Guided Learning for Local Patches to Global Context Awareness in Computational Pathology. arXiv preprint arXiv:2508.19914. 2025 Aug 27.

22. Shahzad, Inzamam, Asif Raza, and Muhammad Waqas. "Medical Image Retrieval using Hybrid Features and Advanced Computational Intelligence Techniques." Spectrum of engineering sciences 3, no. 1 (2025): 22-65
23. Khan, S. U. R., Rehman, H. U., & Bilal, O. (2025). AI-powered cancer diagnosis: classifying viable (live) vs non-viable (dead) cells using transfer learning. Signal, Image and Video Processing, 19(15), 1326.
24. Bilal, O., Hekmat, A., Shahzad, I. et al. Boosting Machine Learning Accuracy for Cardiac Disease Prediction: The Role of Advanced Feature Engineering and Model Optimization. Rev Socionetwork Strat (2025). https://doi.org/10.1007/s12626-025-00190-w
25. Asif Raza, Inzamam Shahzad, Ghazanfar Ali, and Muhammad Hanif Soomro. "Use Transfer Learning VGG16, Inception, and Reset50 to Classify IoT Challenge in Security Domain via Dataset Bench Mark." Journal of Innovative Computing and Emerging Technologies 5, no. 1 (2025).
26. Waqas, Muhammad, Zeshan Khan, Shaheer Anjum, and Muhammad Atif Tahir. "Lung-Wise Tuberculosis Analysis and Automatic CT Report Generation with Hybrid Feature and Ensemble Learning." In CLEF (Working notes), pp. 1-10. 2020.
27. Khan, S. U. R., Asif, S., Zhao, M., Zou, W., & Li, Y. (2025). Optimize brain tumor multiclass classification with manta ray foraging and improved residual block techniques. Multimedia Systems, 31(1), 1-27.
28. Khan, M. A., Khan, S. U. R., Rehman, H. U., Aladhadh, S., & Lin, D. (2025). Robust InceptionV3 with Novel EYENET Weights for Di-EYENET Ocular Surface Imaging Dataset: Integrating Chain Foraging and Cyclone Aging Techniques. International Journal of Computational Intelligence Systems, 18(1), 204.
29. Raza, Asif, Inzamam Shahzad, Muhammad Salahuddin, and Sadia Latif. "Satellite Imagery Employed to Analyze the Extent of Urban Land Transformation in The Punjab District of Pakistan." Journal of Palestine Ahliya University for Research and Studies 4, no. 2 (2025): 17-36.
30. Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Li, X. (2025). Optimized deep learning model for comprehensive medical image analysis across multiple modalities. Neurocomputing, 619, 129182.
31. Raza, Asif, Salahuddin, & Inzamam Shahzad. (2024). Residual Learning Model-Based Classification of COVID-19 Using Chest Radiographs. Spectrum of Engineering Sciences, 2(3), 367–396.
32. Asif Raza, Salahuddin, Ghazanfar Ali, Muhammad Hanif Soomro, Saima Batool, "Analyzing the Impact of Artificial Intelligence on Shaping Consumer Demand in E-Commerce: A Critical Review", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.17, No.5, pp. 42-61, 2025. DOI:10.5815/ijieeb.2025.05.04
33. Khan, Saif Ur Rehman, Asif Raza, Inzamam Shahzad, and Shehzad Khan. "Subcellular Structures Classification in Fluorescence Microscopic Images." In International Conference on Computing & Emerging

Technologies, pp. 271-286. Cham: Springer Nature Switzerland, 2023.

34. Maqsood, H., & Khan, S. U. R. (2025). MeD-3D: A Multimodal Deep Learning Framework for Precise Recurrence Prediction in Clear Cell Renal Cell Carcinoma (ccRCC). arXiv preprint arXiv:2507.07839.
35. Salahuddin, Syed Shahid Abbas, Prince Hamza Shafique, Abdul Manan Razzaq, & Mohsin Ikhlaq. (2024). Enhancing Reliability and Sustainability of Green Communication in Next-Generation Wireless Systems through Energy Harvesting. Journal of Computing & Biomedical Informatics.
36. S. U. R. Khan, A. Raza, I. Shahzad and G. Ali, "Enhancing Concrete and Pavement Crack Prediction through Hierarchical Feature Integration with VGG16 and Triple Classifier Ensemble," 2024 Horizons of Information Technology and Engineering (HITE), Lahore, Pakistan, 2024, pp. 1-6.
37. Mahmood, F., Abbas, K., Raza, A., Khan,M.A., & Khan, P.W. (2019 ). Three Dimensional Agricultural Land Modeling using Unmanned Aerial System (UAS). International Journal of Advanced Computer Science and Applications (IJACSA) [p-ISSN : 2158-107X, e-ISSN : 2156-5570], 10(1).
38. Meeran, M. T., Raza, A., & Din, M. (2018). Advancement in GSM Network to Access Cloud Services. Pakistan Journal of Engineering, Technology & Science [ISSN: 2224-2333], 7(1).
39. Salahuddin, Hussain, M., Hamza Shafique, P., & Abbas, S. S. (2024). Intelligent Melanoma Detection Based On Pigment Network. Kashf Journal Of Multidisciplinary Research, 1(10), 1-14.
40. Khan, S. U. R. (2025). Multi-level feature fusion network for kidney disease detection. Computers in Biology and Medicine, 191, 110214.
41. HUSSAIN, S., Raza, A., MEERAN, M. T., IJAZ, H. M., & JAMALI, S. (2020). Domain Ontology Based Similarity and Analysis in Higher Education. IEEEP New Horizons Journal, 102(1), 11-16.
42. Hekmat, A., Zuping, Z., Bilal, O., & Khan, S. U. R. (2025). Differential evolution-driven optimized ensemble network for brain tumor detection. International Journal of Machine Learning and Cybernetics, 1-26.
43. Raza, A., & Meeran, M. T. (2019). Routine of Encryption in Cognitive Radio Network. Mehran University Research Journal of Engineering and Technology [p-ISSN: 0254-7821, e-ISSN: 2413-7219], 38(3), 609-618.
44. Raza, Asif , Soomro, M. H., Shahzad, I., & Batool, S. (2024). Abstractive Text Summarization for Urdu Language. Journal of Computing & Biomedical Informatics, 7(02).
45. Aslam, M., Salahuddin, Ali, G., & Batool, S. (2024). Assessing The Effects Of Big Data Analytics And Ai On Talent Acquisition And Retention. Kashf Journal Of Multidisciplinary Research, 1(11), 73-84.
46. M. Wajid, M. K. Abid, A. Asif Raza, M. Haroon, and A. Q. Mudasar, “Flood Prediction System Using IOT & Artificial Neural Network”, VFAST trans. softw. eng., vol. 12, no. 1, pp. 210–224, Mar. 2024.
47. N. Mayumu, D. Xiaoheng, P. Mukala, S. U. R. Khan and M. U. Saeed, "Omni-V2X: A Vision-Language Model for Actionable Insights in Vehicle-to-Everything Systems," 2025 International Joint Conference on Neural Networks (IJCNN), Rome, Italy, 2025, pp. 1-8, doi: 10.1109/IJCNN64981.2025.11228491.

48. Bilal, O., Hekmat, A., & Khan, S. U. R. (2025). Automated cervical cancer cell diagnosis via grid search-optimized multi-CNN ensemble networks. Network Modeling Analysis in Health Informatics and Bioinformatics, 14(1), 67.
49. Khan, S. R., Asif Raza, Inzamam Shahzad, & Hafiz Muhammad Ijaz. (2024). Deep transfer CNNs models performance evaluation using unbalanced histopathological breast cancer dataset. Lahore Garrison University Research Journal of Computer Science and Information Technology, 8(1).
50. M. Waqas, Z. Khan, S. U. Ahmed and Asif. Raza, "MIL-Mixer: A Robust Bag Encoding Strategy for Multiple Instance Learning (MIL) using MLP-Mixer," 2023 18th International Conference on Emerging Technologies (ICET), Peshawar, Pakistan, 2023, pp. 22-26.
51. Soomro, M. H., Salahuddin, Irtaza, G., Ali, G., & Batool, S. (2024). Use Image Processing Model To Fruit Quality Detection. Kashf Journal Of Multidisciplinary Research, 1(11), 85-106.
52. Yang, H., Khan, S. U. R., Bilal, O., Chen, C., & Zhao, M. (2025). CEOE-Net: Chaotic Evolution Algorithm-Based Optimized Ensemble Framework Enhanced with Dual-Attention for Alzheimer's Diagnosis. *Computer Modeling in Engineering & Sciences*, *145*(2), 2401.
53. S. ur R. Khan, Asif. Raza, Muhammad Tanveer Meeran, and U. Bilhaj, "Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers", VFAST trans. softw. eng., vol. 11, no. 4, pp. 80–92, Dec. 2023.
54. Chomba, B., Mukala, P., Mayumu, N., & Khan, S. U. R. (2025). DynaKG: Dynamic Knowledge Graph Attention With Learnable Temporal Decay for Recommendation. *IEEE Access*, *13*, 216956-216970.
55. O. Bilal, Asif Raza, S. ur R. Khan, and Ghazanfar Ali, "A Contemporary Secure Microservices Discovery Architecture with Service Tags for Smart City Infrastructures ", VFAST trans. softw. eng., vol. 12, no. 1, pp. 79–92, Mar. 2024
56. Ashraf, M., Jalil, A., Salahuddin & Jamil, F. (2024). Design And Implementation Of Error Isolation In Techno Meter. Kashf Journal Of Multidisciplinary Research, 1(12), 49-66.
57. S. Ur Rehman Khan, O. Bilal, S. Mistry, N. Deb, M. Mahmud and M. Bhuyan, "KDLight: A Lightweight Knowledge Distillation Framework for Medical Image Classification," *2025 International Joint Conference on Neural Networks (IJCNN)*, Rome, Italy, 2025, pp. 1-8, doi: 10.1109/IJCNN64981.2025.11228615.
58. O. Bilal, S. Ur Rehman Khan, S. Mistry, N. Deb, M. Mahmud and M. Bhuyan, "Towards Efficient Pruning and Multi-Scale Feature Transformations to Uncover Medical Diseases," *2025 International Joint Conference on Neural Networks (IJCNN)*, Rome, Italy, 2025, pp. 1-8, doi: 10.1109/IJCNN64981.2025.11229047.
59. N. Mayumu, X. Deng, A. Bagula, S. u. R. Khan and P. Mukala, "V2X-JEPA: Self-Supervised Multi-Agent Joint Embedding Predictive Architecture for Robust Vehicle-to-Everything Perception," in *IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2026.3660030.
60. Shahzad, Inzamam, Asif Raza, Hasaan Maqsood, Saif Ur Rehman Khan, and Ghazanfar Ali. "Towards Robust Breast Cancer Diagnosis: A Hybrid Deep Learning Ensemble Framework." In 2025 Horizons of

Information Technology and Engineering (HITE), pp. 1-6. IEEE, 2025.

61. Ishfaque, M., Khan, S. U. R., & Lou, Y. L. (2026). Towards efficient dam inspection: crack detection via chirplet transform feature and a pruned VGG16 architecture. *Memetic Computing*, *18*(1), 9.
62. Sohail, Muhammad Faisal, Hassan Raza Khan, And Muhammad Tanveer Meeran. "Convolutional Neural Network Architectures For Robust Image Feature Representation In Retrieval Systems." Kashf Journal Of Multidisciplinary Research 3, No. 02 (2026): 109-139.
63. Bilal, O., Hekmat, A., Khan, S. U. R., Raza, A., & Ali, G. (2025, December). MS-STO-Net: A Multi-Scale State Transition Optimization-Based Ensemble Network for Accurate White Blood Cell Classification. In *2025 27th International Multitopic Conference (INMIC)* (pp. 1-6). IEEE.