# ANALYSIS OF CARDIOVASCULAR DISEASE USING MACHINE LEARNING

**Asif Ahmad** MTech Student Department of Computer Science and Information Technology, (New Building) Maulana Azad National Urdu University, Gachibowli, Hyderabad,Telangana,Pin –500032
**Mohammad Islam** Assistant Professor, Department of Computer Science and Information Technology, (New Building) Maulana Azad National Urdu University, Gachibowli, Hyderabad,Telangana,Pin -500032

**Abstract-** Machine learning is adaptational method of learning within the medical trade throughout the world. Machine Learning play a necessary role in predicting presence/absence of heart diseases. Heart pumps the blood and delivers it to all the body's organs. Data analysis is a crucial procedure and method in the medical industry that helps anticipate heart status [5], as well as other information and avoids various diseases in hospitals. The main goal of the study paper is to forecast a patient's risk of developing cardiac disease or heart disease using a machine learning-based method. Machine learning methods for predicting cardiac disease include artificial neural networks (ANN) [9], decision trees [3], K-Nearest Neighbours (KNN), naïve bays [14], and support vector machines (SVM). This research's ultimate finding—after studying machine learning—is that the random forest method is the most effective one for the data. The dataset came from UC Irvine.

**Keywords:** Machine learning, Prediction of cardiovascular Disease, K-Neighbor Classifier, Care Vector Classifier, Decision Tree Classifier, Random Forest Classifier.

## 1. Introduction

The heart plays a significant role in the human body. It allows for the distribution of blood throughout the body. The prevalence of numerous heart illnesses has increased as a result of lifestyle modifications, work-related stress, and poor diets [14]. One of the biggest causes of death in the globe is heart disease. Accurate prediction and prediction of heart disorders are essential. Medical organizations around the globe gather information on a variety of health-related topics. To make sense of this data, a variety of machine learning approaches can be applied. However, the data that has been gathered is enormous and usually noisy. These datasets are confusing and are simple to analyze using different machine learning techniques.

### 1.1 Dimensionality Reduction

Dimensionality refers to the process of selecting a mathematical representation that enables the majority, but not all, of the distribution of a particular data set to be linked in order to contain just the most relevant information. Despite the fact that the data examined for a job or issue may include a large number of characteristics or dimensions, not all of them have an impact on the outcome. Many attributes or functions can have an impact on computer complexity or cause overlap, resulting in poor results. As a result, reducing dimensionality is a critical step in the development of a model. Dimension reduction typically takes two categories as one as feature extraction and other as feature selection [1,4.1].

### Extraction of Feature

It is in this situation that a new feature set is created by deriving from the existing feature set. The extraction feature includes a transformation function that is separate from the extraction feature. In most cases, this change is irreversible. It should be noted that throughout this transition, no valuable information is lost. The major component analysis method is a linear transformation technique that is widely used. Find the orthogonal directions as well as the direction that has the greatest variance in the feature's area by using the coordinate system. It is a global algorithm that does the most effective refactoring available today.

**1.2 heart disease prediction**

The size of the data collection is decreased by machine learning algorithms using the dimension reduction technique. It is the earliest and most fundamental technique used to screen the important data sets that have a major impact on disease. In the stage that follows this prediction, the data has finally been modeled for prediction [2,5,6,10].

Omitting the most important information unresolved value problems in place of missing numbers, use the mean or median

Create two groups by dividing the data sets.

A train and a test data set, respectively. Prefer a ratio of 75:25 for the train and test data sets.

In order to analyze the data sets, use logistic regression.

See the accuracy and precision the optimal algorithm to employ

**2. Background**

[1] In this article, they suggested a fluid logic system interval type 2 and a rough, fixed-attribute reduction-based cardiac diagnostic system (IT2FLS). To deal with the difficulties and uncertainties posed by high-dimensional datasets, lumpy sets of attribute reduction and IT2FLS integration are used. The hybrid learning process used by IT2FLS comprised parameterization utilizing chaotic fire-fly algorithms, fuzzy c-means clustering, and hybrid genetic algorithms [2]. This learning process requires a lot of computer time, especially when working with huge datasets. Try using chaotic firefly algorithms to reduce a list of qualities to discover the optimal reduction. This enhances IT2FLS presentation while reducing computational complexity. Preliminary results show that Nave Bayers, vector machines, and artificial neural networks perform better than traditional machine learning techniques. The suggested model can therefore be applied as a decision support system [3] for the detection of heart disorders.

[2] Medical mistakes are frequently expensive and harmful. They murder a lot of individuals each year all across the world. Systems for clinical decision assistance offer the potential to increase patient safety while lowering medical errors. The identification and treatment of cardiac disease is one of the most important features of using this device. Apply categorization methods from data mining to various heart-related issues. This study aims to develop a data mining clustering-based cardiac disease prediction system. A fitness care system is one that is data-rich but knowledge-deficient. In order to improve disease prediction, knowledge will be extracted from these healthcare data. Clinical expert systems are now frequently using data mining technology to forecast various diseases. These tools have uncovered previously unidentified links and patterns in medical data. As a result, it is thought to be a significant task to try to assist the diagnostic process with the expertise and experience of multiple specialists, as well as clinical screening data from patients in the database. Unfortunately, the healthcare sector gathers a lot of information on heart illness that can't be accurately diagnosed in order to find out what's going on.

[3] This study aims to identify heart illness using computer techniques like genetic algorithms and fuzzy logic. This technology aids medical professionals in automating the diagnosis and treatment of cardiac disease. For the purpose of detecting cardiac problems, they develop a hybrid gene. An ideal answer to the function selection issue for random searches is offered by genetic algorithms. The creation of the diagnostic system utilizing a fluid inference system is aided by the selection of the suitable features from the data set. Sample data is used to build fuzzy system rules. Use genetic algorithms to choose significant and pertinent subsets of rules throughout the rule collection. In order to effectively anticipate cardiac disease in patients, the suggested study makes use of genetic algorithms and dynamic thinking systems. Exercise-induced angina, exercise-induced resting suppression of ST (oldpeak), cholesterol (chol), maximum heart rate (thalach), a significant number of vessels, and exercise-induced angina are the

features that were selected (ca). The centroid approach and the FUZZY Gaussian membership feature can be used to enhance system performance. Performance measures like accuracy, specificity, sensitivity, and a confusion matrix were utilized to assess the work's effectiveness. The classification accuracy of the layered k-fold approach was 86%, and its specificity and sensitivity metrics were 0.90 and 0.80, respectively. The cardiac illness dataset in the UCI machine learning repository now only has 7 attributes, down from 13 previously. The proposed approach boosts accuracy by 1.54% when compared to the current system. The proposed model is denoted as the GAFL model for efficient cardiovascular prediction. Modeling is straightforward and gives clinicians in hospitals and medical facilities a practical choice. The proposed approach boosts accuracy by 1.54% when compared to the current system. The proposed model is denoted as the GAFL model for efficient cardiovascular prediction. Modeling is straightforward and gives clinicians in hospitals and medical facilities a practical choice.

[4] Due to the tremendous workload, heart condition prediction is a serious concern, and heart disease prediction has emerged as the most worrisome problem. The condition is particularly challenging to diagnose. The problem is that no useful information was used to retrieve the data. In order to extract useful data, data mining techniques are employed. Forecasting cardiac illness does not involve the decision tree or ID3.. Heart disease prediction is a topic that many scientists and professionals are familiar with and are skilled at using. A decision tree is built to forecast cardiac illness in order to resolve this issue. In this work, the data was pre-processed, and cardiovascular illnesses were predicted using a decision tree algorithm using ID3.

[5] In the world, heart disease is the main cause of death. Predicting the result of a unwellness may be a tough task. methoding} generates automatic diagnostic rules and assists specialists in increasing the responsibility of the diagnostic process. Researchers use a spread of information mining techniques to help medical professionals in predicting vessel diseases. Accidental Forest may be a medical data algorithmic program that's each integrated and precise. Chi-square operate choice metrics area unit want to assess and establish the relationships between variables. This paper proposes a classification model for predicting cardiopathy supported random forest [4], a Chi-square methodology, Associate in Nursingd an transmitted algorithmic program. The experimental results show that their methodology outperforms different classification procedures in terms of classification preciseness, and medical professionals will with success use the projected model within the prediction of cardiopathy.

[6] cardiovascular disease is one of the primary causes of morbidity and mortality. It is crucial to identify cardiovascular illnesses, but it must be done so with great care and effectiveness, and finding the right automation is difficult to work. A doctor is incomparable to anyone. In many locations, doctors with easy access to professional skills are hard to come by. It is difficult for all doctors to have the same expertise in all sub-professionals. Automated systems reduce the cost of medical diagnostics while improving patient care. In this work, a plan was developed to efficiently find guidelines for estimating patient risk levels based on specified health data. The requirements of the user might determine the rule's priority. The system is more capable of accurately predicting the risk of heart disease, as evidenced by its performance as assessed by classification accuracy.

[7] One of the main causes of death worldwide is heart disease. The ability to forecast upset may be a key issue in the realm of clinical information analysis. The enormous amount of data produced by the aid industry has been demonstrated to support machine learning (ML) in decision-making and prediction. they need seen the utilization of milliliter technology all told areas of the net of Things in recent developments (IoT). varied studies solely give information on vessel prediction exploitation milliliter technology. This document proposes the utilization of machine learning techniques to notice vital options so as to enhance

the prognosticative accuracy of vessel diseases. it absolutely was introduced a prognosticative model with varied mixtures of characteristics and several other well-known classification techniques. A random biological science and linear model (HRFLM) will improve performance by 88:7%.

[8] Based on the converting coefficients generated by the point propagation function of the Chebyshev orthogonal polynomials, they suggest a new edge detector. The Prewitt and Roberts-like edges were detected by the edge detector. Depending on the conversion factor, sensitive to a variable parameter that can be calculated. As an illustration, the brain can be eliminated from magnetic resonance imaging (MRI) of a human head using an edge detector.

[9] One of the most significant illnesses that affect humans, heart disease has a terrible effect on people's life. The heart cannot pump enough blood to the body's tissues during a heart attack. In order to prevent and cure heart failure, it is essential to get an accurate and prompt diagnosis of heart disease. The diagnosis of heart disease based on a conventional medical history is viewed as being unreliable in many aspects. Healthy individuals and cardiac disease patients can be distinguished using non-invasive techniques like machine learning. A cardiac predictive diagnostic system that automatically learnt from datasets on heart illness was developed as part of the proposed study. Each classifier calculates the optimistic curve as well as the curve's area under it. All classifiers, function selection algorithms, preprocessing techniques, validation techniques, and methodologies for classifier performance evaluation were described in this study. To evaluate the 0e performance of the suggested system, a comprehensive and condensed set of functions is used. The execution and accuracy of categorization are both impacted by reduced 0e function. The 0e machine-based decision-making system enables doctors to accurately diagnose cardiac patients.

[10] In the history of medical data, data mining techniques have been most widely used and investigated, and the prognosis of heart disease has shown to be crucial in medicine. It was discovered that medical history data were heterogeneous, and several different types of data needed to be analyzed to foretell the patient's cardiac condition. For the purpose of predicting heart disease patients, many data mining techniques have been used. Data mining techniques do not, however, remove data ambiguity. In order to remove uncertainty, ambiguity was tried in the measurement data. Measurements are paired with membership features, which are designed to eliminate uncertainty. Additionally, an effort was made to group patients according to their medical characteristics. To categories data among groups, the minimum distance of the K-NN classifier is combined. This demonstrates how the K-NN fuzzy classification is superior to other technological parameter classifiers in terms of appropriateness.

[11] One of the leading causes of mortality and morbidity in the US is heart disease. The likelihood that a patient would acquire cardiac disease can be predicted using data mining techniques. This study looked at several data mining techniques' forecasts for heart disease. In this study, data mining techniques for estimating the risk of cardiovascular disease are compared. After performing feature testing, the C5.0, neural network, support vector machine (SVM), K-Nearest Neighborhood (KNN), and logistical regression models were built and validated. The decision tree C5.0 can locate the best accurate model with a 93.02 percent accuracy rate. KNN, SVM, and neural networks make up 88.37, 86.05, and 80.23 percent of the total, respectively.
The outputs of the decision tree are simple to use and comprehend. The decision tree findings are straightforward to comprehend and put into practice, and various physicians can quickly pick up on the guidelines.

[12] The amount of data produced by treatment facilities, therapy centers, and medical organizations is so great that it is not fully exploited. The medical system is "data-rich," yet it is not enough. There are no

reliable analytical techniques for finding relationships and trends in medical data. The data mining approach is helpful in this case. various data mining approaches can therefore be applied. This White Paper's goal is to provide several knowledge abstraction methods that utilize current prediction data mining methods. This white paper explores decision-tab algorithms for medical datasets, neural systems, and Naïve Bayes [7,15] data mining techniques.

## 3. Algorithms Used

### Random Forest Classifier Methodology

Random Forest is a supervised machine learning [8] method that is also used in other fields. It is possible to apply this method for both regression and classification problems; however, it is more successful when used for classification tasks. Random forest [4] approaches, as their name suggests, evaluate a large number of different decision trees before providing a final outcome. A collection of decision trees [6] is what it comes down to becoming as a consequence of this. This approach is predicated on the idea that additional trees will eventually arrive at the right choice in the situation. In classification [11], a voting system is utilized to select the category, and in regression, the average of all decision tree outputs is achieved by averaging the outputs from all decision trees. It is suitable for big data sets with a high degree of dimensionality.

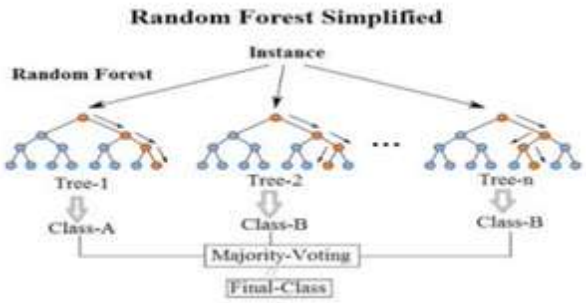



**Fig:** Figure shows the Random Forest

### Support Vector Machine

support in supervised machine learning [8, 13] (with pre-defined target variables), vector is a common method for classifying and predicting that may be used to classify and predict. Hyperplanes that can be categorized should be sought for in the feature room. Training data points are represented as function space points in an SVM model, and the model maps them in order to distinguish as many points as possible from various classes. On the margin side of the same space, Mapped and sorted test data points.

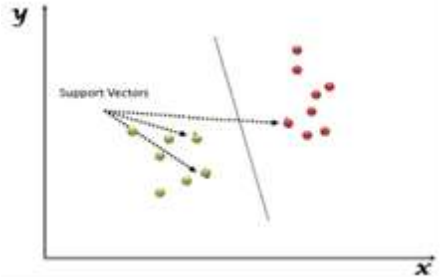



**Fig:** A presentation of SVM

### Decision Tree

An example of a supervised learning algorithm is a decision tree that may be used to acquire new skills. This method is mostly used in the solution of classification issues [11]. With continuous and categorical characteristics, you can simply perform. Because the most significant predictors are used to divide the population, this method divides the population into two or more comparable groups. Before anything else,

the decision tree method estimates the entropy for each characteristic. The data set is divided by highest information gain or lowest entropy for the variable or predictor [3,5].

**Logistic Regression**

Regression can be defined as the measurement and study of the connection between one or more independent and dependent variables. Logistic regression and linear regression are the two forms of regression. Logistic regression is widened by applying linear regression. Since the response variables are discrete, linear regression cannot adequately represent them. They are typically employed to estimate dependent variables that are binary or multiclass.

**K – Nearest Neighbor**

In 1951, Hodges and colleagues developed a non-parametric technique for pattern categorization that was widely used. This is referred to as the K-Nearest Neighbor Rule, which is well-known. The technology of K-Nearest Neighbors is one of the most fundamental and successful classification methods [11] available. It is used for classification tasks that don't build assumptions regarding the information and customarily have very little or no previous data regarding the distribution of knowledge. This technique determines the k nearest information points within the coaching set that ar nearest to the out-of-stock information points within the coaching set and determines the common price of the information points that are nearest to the information points not out there within the coaching set. Specifically, within the case of ranking, the closest neighbor technique essentially creates a majority vote among the K most comparable instances of a specific "invisible" observation that exists. The similarity between 2 information points is outlined by the space metric between the 2 information points. a well-liked selection is that the Euclidian distance given by

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + \ldots + (x_n - x_n')^2}$$

**4 Flow Chart of Execution**


**4.1 Data Source**

This paper utilized a dataset from the University of California, Irvine Machine Learning [8,13] repository for my investigation. In it, you will get blood pressure, chest pain, electrocardiogram result, etc., and so on. As part of our study, we utilized five algorithms to determine the causes of heart disease [5, 11] and to build a model with the highest level of accuracy feasible.

**4.2 Data Pre-processing**

The data contains a sizable number of missing and erroneous numbers. To prevent these issues and generate more precise projections, these data have been pre-processed. Illustration 1 depicts the sequential chart of our proposed model, which we believe is accurate. Cleaning the acquired data is usually required due to the presence of noise and missing values in the data. This data must be cleansed of noise and any missing values must be filled in to ensure that the data is full in order to provide an accurate and effective result. When you transform data from one format to another, you are making it more comprehensible for the person who is looking at it. It consists of activities like as smoothing, standardization, and aggregation.

Integration Data may not be obtained from a single source, but from a number of different sources, and it must be integrated before it can be processed. The data obtained during the reduction process is complicated and must be structured in order to provide appropriate results. The data is then categorized and divided into two sets: a training data set and a test data set, each of which is subjected to a variety of algorithms in order to provide accuracy score results.

**Proposed Steps for Data Modeling**

Step 1 is to identify key characteristics in heart data sets. For statistical analysis, the property with the smallest and largest data sets is chosen.
STEP 2: Use statistical analysis to determine whether the data are normal.
STEP 3: Assess the mean and median procedures for filling in the missing values.
STEP 4: Use the means and median of the data sets to fill in the missing values.
STEP 5: Use a 70:30 ratio to divide the data for the test and train analyses.
STEP 6: Use the train data sets to run the learning algorithm.
STEP 7: Utilize the test data sets to evaluate the precision.

**5.Attributes**

This study seeks to forecast the likelihood of having heart disease as a likely cause of computerized prediction of heart disease [15] that is useful in the medical field for doctors and patients [5] by using statistical methods. In order to achieve this goal, we have addressed the application of different machine learning algorithms [13] to the data set, as well as the importance of dataset analysis in this research article. In addition, this study illustrates which characteristics contribute more than others to the prediction of greater accuracy in the future. In certain cases, this may save the patient the cost of many trials since not all of the characteristics may contribute in such a significant way to predicting the result [5].

| Sr. no. | Attribute | Representative icon | Details |
|---|---|---|---|
| 1 | Age | Age | Patients age, in years |
| 2 | Sex | Sex | 0 = female; 1 = male |
| 3 | Chest pain | Cp | 4 types of chest pain (1—typical angina; 2—atypical angina; 3—non-anginal pain; 4—asymptomatic) |
| 4 | Rest blood pressure | Trestbps | Resting systolic blood pressure (in mm Hg on admission to the hospital) |
| 5 | Serum cholesterol | Chol | Serum cholesterol in mg/dl |
| 6 | Fasting blood sugar | Fbs | Fasting blood sugar > 120 mg/dl (0—false; 1—true) |
| 7 | Rest electrocardiograph | Restecg | 0—normal; 1—having ST-T wave abnormality; 2—left ventricular hypertrophy |
| 8 | MaxHeart rate | Thalch | Maximum heart rate achieved |
| 9 | Exercise-induced angina | Exang | Exercise-induced angina (0—no; 1—yes) |
| 10 | ST depression | Oldpeak | ST depression induced by exercise relative to rest |
| 11 | Slope | Slope | slope of the peak exercise ST segment (1—upsloping; 2—flat; 3—down sloping) |
| 12 | No. of vessels | Ca | No. of major vessels (0–3) colored by fluoroscopy |
| 13 | Thalassemia | Thal | Defect types; 3—normal; 6—fixed defect; 7—reversible defect |
| 14 | Num(class attribute) | Class | diagnosis of heart disease status (0—nil risk; 1—low risk; 2—potential risk; 3—high risk; 4—very high risk) |

Characteristics and Predictor: Thirteen characteristics (X) define our Predictor (Y, positive or negative diagnosis of heart disease) [5].

**6. Steps for Modelling**

Step1: The data preparation is to sketch up an outline of the data.
Outliers should be identified and eliminated.

Identify and handle the data that is missing.
Using appropriate normalizing methods to get the desired results
Take the place of the mean and median.
}
Step 2: Choosing the Right Model
{
Investigating the value of data (classes)
Soft-learning learning Algorithm Sorting is a learning algorithm that is based on soft-learning learning.
}
Step 3: Python Programming for Model Implementation
{
Incorporating all models into a single Python programmed is called import data.
}
Step 4: Performance Evaluation and Classification
{
Determine Accuracy by calculating it using the "Performance" operator and evaluating the outcome via the accuracy measure
}
Step 5: Evaluation of the results
{
evaluating and comparing each model's accuracy comparing the outcomes with all of the available soft learning algorithms
Calculate the ultimate output of each algorithm that has been proposed
Find the most suitable candidate among all of them.
}

**Classification Rate/ Accuracy:**

Classification Rate or Accuracy is given by the relation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Classification Rate or Accuracy is given by the relation:

There are, however, issues with the veracity of the data. It is predicated on the assumption of equal costs for both types of mistakes. Depending on the issue, a 99 percent accuracy rate may be great, good, average, bad, or awful depending on the accuracy rate.

**Recall**

Recall is the ratio of the total number of positive cases that were correctly identified over the total number of positive cases when it comes to positive cases. An elevated rate of reminders indicates that the category has been accurately classified (the number of FNs is small).

$$\textit{Recall} = \frac{TP}{TP + FN}$$

**F-measure**

It is beneficial to have a single measurement that reflects both precision and recall, since we have two measurements (precision and recall). We calculate Fmeasure using harmonic averages instead of arithmetic averages because it penalizes extreme values more severely. The largest Precision or Recall value will never be closer to FMeasure than the smallest Precision or Recall value.

The data set was evaluated using various classifiers for accuracy, precision, and F1-Score in the prediction of heart disease [15]. As find in the above table all eleven classifier has been executed through machine learning algorithms. The result has been presented in above table could easy been analyzed than the level of accuracy with their precision value. The above result stated that the highest accuracy provider among them is SVM Prediction, Gaussian NB and Linear Discriminant Analysis which is 86.89. The accuracy is most accepted things that could be considered for selecting the best classifier for the prediction and constructing the frame work.

Further an GUI also constructed in MATLAB for real time prediction. The technique behind used is sampling based on Fuzzy categorization for finding the threshold value.

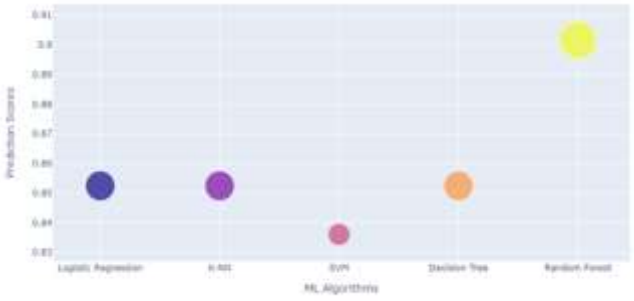
**Fig:** A GUI (MATLAB) for User input

**Fig:** The simulation popup predicts the Heart

The Constructed GUI is as above which used as the input for user to get the prediction of Heart on the basis of the diagnosis data.

**Table 1:** Comparative table of Machine learning algorithms on Heart data set

| **Algorithms** | **Logistic Regression** | **KNN Prediction** | **SVM Prediction** | **Decision Tree** | **Random Forest** |
|---|---|---|---|---|---|
| **Accuracy** | 0.852 | 0.852 | 0.868 | 0.852 | **0.901** |
| **precision** | 0.8823 | 0.8823 | 0.8823 | 0.8823 | 0.8823 |
| **f1-score** | 0.909 | 0.909 | 0.909 | 0.909 | 0.909 |



## Conclusion and Future work

This article discussed a few reliable methods for anticipating cardiac disease and evaluated the accuracy of the organization method based on the selected classification algorithm. A significant test in data mining and machine learning is the development of accurate and computationally efficient classifiers for medical applications. In the future, it will be very challenging for vast populations all over the world to treat a significant number of patients. Modernization has a considerable impact on the lifestyle of the urban people in addition to lifestyle changes. In this situation, a computerized system that aids doctors in disease prediction is necessary. The success and best prediction analysis of several machine learning approaches are shown in this study.

## References

1. Long, N. C., Meesad, P., & Unger, H. (2015). A highly accurate firefly-based algorithm for heart disease prediction. *Expert Systems with Applications*, *42*(21), 8221-8231.
2. Santhanam, T., & Ephzibah, E. P. (2015). Heart disease prediction using hybrid genetic fuzzy model. *Indian Journal of Science and Technology*, *8*(9), 797.
3. Javed, S., Javed, H., Saddique, A., & Rafiq, B. (2018). Human Heart Disease Prediction System Using Data Mining Techniques. *Sir Syed Research Journal of Engineering & Technology*, *8*(II).
4. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. *Journal of Network and Innovative Computing*, *4*(2016), 175-184.
5. Saxena, K., & Sharma, R. (2015, May). Efficient heart disease prediction system using decision tree. In *International Conference on Computing, Communication & Automation* (pp. 72-77). IEEE.
6. Saxena, K., & Sharma, R. (2015, May). Efficient heart disease prediction system using decision tree. In *International Conference on Computing, Communication & Automation* (pp. 72-77). IEEE.
7. Sharmila, S., & Gandhi, M. I. (2017). Heart Disease Prediction Using Data Mining Techniques-Comparative Study. *Computational Methods, Communication Techniques and Informatics*, 351.
8. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *2018*.
9. Abdar, M., Kalhori, S. R. N., Sutikno, T., Subroto, I. M. I., & Arji, G. (2015). Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *International Journal of Electrical & Computer Engineering (2088-8708)*, *5*(6).
10. Shinde, A., Kale, S., Samant, R., Naik, A., & Ghorpade, S. (2017). Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network. *International Journal of Computer Applications*, *166*(7), 32-36.
11. Gandhi, M., & Singh, S. N. (2015, February). Predictions in heart disease using techniques of data mining. In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*(pp. 520-525). IEEE.
12. A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, (2015.) "Effective diagnosis and monitoring of heart disease", *International Journal of Software Engineering and Its Applications*, Vol.9, No.1, pp. 143-156.
13. G. Parthiban and S. K. Srivatsa, (2012)"Applying machine learning methods in diagnosing heart disease for diabetic patients", *International Journal of Applied Information Systems*, Vol.3, No.7, pp.2249-0868.
14. K. Vembandasamy, R. Sasipriya, and E. Deepa, (2015)"Heart Diseases Detection Using Naive Bayes Algorithm", *IJISET-International Journal of Innovative Science, Engineering & Technology*, Vol.2, pp.441-444.
15. Patel J, Tejal Upadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Dis. 2015;7(1):129–37.