I. J. Engineering and Manufacturing, 2026, 2, 214-225
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.02.14

A Systematic Approach for Heart Disease Analysis using Machine Learning Algorithms

Mahmood Ali Mirza*
Department of Computer Science and Engineering , Krishna University College of Engineering and Technology, Krishna University, Rudravaram, Andhra Pradesh, India.
E-mail: alimirza.md@gmail.com
ORCID ID: https://orcid.org/0000-0003-3071-1674
*Corresponding Author

Jaddu Lohith Kumar
Department of Computer Science and Engineering , Krishna University College of Engineering and Technology, Krishna University, Rudravaram, Andhra Pradesh, India.
ORCID ID: https://orcid.org/0009-0004-7773-0236

Gandavarapu Mohan Sai
Department of Computer Science and Engineering , Krishna University College of Engineering and Technology, Krishna University, Rudravaram, Andhra Pradesh, India.
ORCID ID: https://orcid.org/0009-0003-4001-5725

Kongora Venkat Chowdary
Department of Computer Science and Engineering , Krishna University College of Engineering and Technology, Krishna University, Rudravaram, Andhra Pradesh, India.
ORCID ID: https://orcid.org/my-orcid?orcid=0009-0009-1795-4218

Karnaati Kranthi Kiran
Department of Computer Science and Engineering , Krishna University College of Engineering and Technology, Krishna University, Rudravaram, Andhra Pradesh, India.
ORCID ID: https://orcid.org/0009-0009-3397-1480

Faheem Ali Mirza
Department of Information Technology, SIR C R Reddy College of Engineering, Eluru, Andhra Pradesh, India.
E-mail: faheemalimirza07@gmail.com
ORCID ID: https://orcid.org/0009-0004-8117-2812

Received: 03 April, 2025; Revised: 23 May, 2025; Accepted: 14 September, 2025; Published: 08 April, 2026

Abstract: The number of heart disease patients has significantly increased in recent years, and heart illness is linked to a high death rate. Furthermore, as technology advanced, several sophisticated devices were created to assist patients in measuring their health at home and estimating their risk of developing heart disease. Using six machine learning models, the study seeks to determine how accurate self-measured physical health indicators are at predicting heart disease when compared to all indicators assessed by medical professionals. Logistics Regression, K Nearest Neighbors, Support Vector Model, Decision Tree, Random Forest, and Gradient Boosting Classifier were among the six models employed to forecast heart disease. Twelve different test findings and 1189 patients' heart disease risks are included in the database utilized for the study. While the metrics contains six outcomes that could be tested, the all metrics contained all twelve test results. The accuracy score and false negative rate were calculated for each of the five models that were built for the all metrics.The findings demonstrated that in all five models, all metrics had greater accuracy scores than existing metrics. For five machine learning models, all metrics had false negative rates that were either lower or equivalent to that of existing metrics. The results showed that all physical indicators were more accurate in predicting patients' risk of heart disease than metrics measured physical health indicators. Therefore, all physical health indicators are preferable for assessing the risk for cardiac illnesses in the absence of future development of indicators.

Index Terms: Data Mining, Classification, Heart disease, Regression, K Nearest Neighbors, Support Vector Model, Decision tree, and Random Forest, Gradient Boosting Classifier.

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

Heart disease is generally regarded as a direct threat to human life and health because it is brought on by abnormal heart and blood vessel abnormalities. It is one of the major illnesses that many middle-aged and older people are experiencing irreversible consequences from, with a high risk of deadly complications [1,2]. According to Makino, among those 65 years of age or older, the absolute risk of cardiovascular heart disease is linked to both disability and death [3.4]. According to estimates from the World Health Organization (WHO), 17.7 million deaths in 2015 were attributed to cardiovascular illnesses, which accounted for one-third of all fatalities in that year [5].

One of the two leading causes of death in Australia, according to the Australian Bureau of Statistics, was heart disease [6.7]. Since heart disease has such a detrimental impact on human health, a lot of research has been done on how it develops and how to prevent and lower its incidence in a timely and effective manner. Furthermore, it is advised to employ advanced equipment to identify any heart risks beforehand in order to prevent the negative impacts of heart disease.

Numerous tests, such as blood tests, echocardiography, chest X-rays, magnetic resonance imaging (MRI), electrocardiograms, physical examinations, and exercise stress tests, can currently be performed by qualified health organizations. These tests give doctors important information for diagnosing patients and determining their risk of heart failure [8].

Heart failure has multiple risk factors, each of which correlates to a separate test index. To identify the possible characteristics of a heart attack, a substantial quantity of pertinent research has been conducted. Heart disease is dependent on age, sex, smoking, high blood pressure, and diabetes [9]. Numerous people contend that indicators such as age, blood pressure, and total cholesterol are crucial for predicting coronary heart disease [10,11]. At COVID-19, sex differences are thought to have a significant impact on traditional cardiovascular risk variables [12].

Additionally, heart rate is a strong predictor of a patient's chance of having a heart attack [13]. Depending on whether the indications could be assessed at home, the characteristics of heart disease may be roughly classified into two categories. In order to improve future heart disease testing, it is thought to be beneficial to compare the accuracy of markers assessed at home and in hospitals.

In order to determine the connection between heart illnesses and patients' medical circumstances, computational technology and statistical methods have gained popularity [10], [14,15]. They can lower the chance of dying from a heart attack by assisting in the early prediction of the patient's possible risk of heart disease depending on their underlying physical state. Numerous computer-based statistical techniques have been used to forecast cardiac attacks [16]. SVM has been widely used as a classification technique to forecast heart attacks because of its high accuracy [17].

In order to evaluate heart failure and achieve compromise outcomes, they employed logistic regression and the k-NN method [18]. By altering feature selection, Random Forest has been used to obtain the highest accuracy of 82.18% [19]. Researchers and medical professionals can make better decisions regarding heart disease because these algorithms have been shown to accurately predict the risk of heart disease.

2. Literature Review

Tyagi et al. [20] proposed a novel method to improve the accuracy of heart disease prediction by strategically using machine learning techniques to find important information. By analyzing different feature combinations and applying popular classification methods such as decision trees, logistic regression, random forest, K-nearest neighbors (KNN), gradient boosting machines (GBM), XG Boost, and multilayer perceptrons, this study seeks to develop a trustworthy prediction model. By conducting this thorough investigation, we hope to improve cardiac health prediction, which could lead to better patient outcomes and less heart disease's impact on international healthcare systems. In order to accurately forecast the occurrence of heart disease in a patient, Mishra et al. [21] have examined various ensemble methods (Bagging [28], MaxVoting [29], Boosting [30], Random Forest [31]) as well as K-Nearest Neighbor Method, Naïve Bayes, Logistic Regression, Support Vector Machine, and Decision Tree. According to the experimental findings, the maximum accuracy was attained using the ensemble method, or Max Voting.

[22] Ekle et al. In addition to proposing a Dietary Approach to Stop Hypertension (DASH) lifestyle change therapy recommendation system design for heart disease, the study will conduct a thorough comparative analysis of the performance of various supervised learning machine learning models for heart disease prediction. The authors of this study compared 61 papers from the PubMed and Google Scholar databases that employed multiple supervised learning machine learning algorithms for the prediction of heart disease. The suggested system blueprint was created using a content-based filtering suggestion technique. The Voting Ensembles Classifier (VEC) algorithm had the highest accuracy in comparison. This depends on the notion that, even while individual model could marginally over fit or under fit the data, when combined, their errors can cancel each other out to yield more reliable and accurate forecasts.

Additionally, the total effectiveness of healthcare administration can be enhanced by VEC's more accurate forecasts. Finally, the design of the suggested dietary therapy recommendation system for heart disease was presented in this study. In addition to providing a comprehensive overview of the relative effectiveness of various supervised learning machine learning algorithms for the prediction of heart disease, this study suggests a framework for a

recommendation system for dietary lifestyle change therapy. When completely implemented, the suggested system can serve as a dietary therapy support tool for cardiologists, and the comparison performance data can help researchers select an appropriate machine learning method for their studies.

According to Kuldeep Vayadande et al., [23] WHO data, cardiac disorders cause around 1 crore 20 lakh deaths annually. Heart disease and cardiovascular disease are extremely harmful and have a significant global influence, as they have had a significant impact on the medical industry in previous years. Although it takes a lot of time and knowledge to forecast heart problems or CD, it is also impossible to monitor patients around-the-clock.The diagnosis and treatment of heart disease can be extremely challenging, particularly in underdeveloped or impoverished nations. Additionally, a person may die as a result of inadequate medical care or a delayed diagnosis of a disease. Researchers constantly use the medical industry's vast data to create new science and technologies that will reduce the number of heart disease-related deaths. The several machine learning methods for predicting heart disease from datasets are the main topic of this research.

According to study by Mallikarjunamallu et al. [24], there are a few issues with current methods that have not yet been fixed when working big datasets, which eventually lessens their actual utility and effectiveness. To improve data-driven decision-making for the identification of cardiac disease, ML and DL techniques are applied. Thirty articles about heart disease diagnosis were selected for content analysis of the metadata of ninety-three articles. The performance of the models was the main focus of the study, along with other concerns like the interpretability and explicability of machine learning and deep learning.

The focus of G. Sakthipriya et al. [25] is on the pertinent risk factors that lead to heart disorders. These risk factors are divided into two groups. The first category consists of the patient's age, sex, and family history. The second category includes risk factors relating to the patient's lifestyle. Among these include high blood pressure, smoking, high cholesterol, and sedentary lifestyles. To forecast the risk level of cardiac disorders, we have employed algorithms such as Random Forest, Support Vector Classifier, and Logistic Regression. We have also conducted a comparison analysis of these three methods.

Prediction accuracy, a performance criterion employed by Bhavekar et al. [26] to predict cardiac disease, is still below ideal. A person can protect himself from life-threatening situations with the aid of an accurate HDP, but an incorrect prediction can also be lethal. Several Deep Learning (DL), Machine Learning (ML), and optimization-based HDP strategies are covered in this review work in order to address these problems. In order to assist experts and the healthcare business in predicting cardiac disease, numerous researchers have recently begun using various DL and ML algorithms. It also covered the performance study of several optimization-based algorithms. Thus, the optimization-based HDP method may help physicians anticipate the onset of heart disease and provide appropriate treatment, according to this review report.

To solve this problem, Amrit Singh et al. [27] developed a unique diagnostic system that performs exceptionally well on both training and testing data and exhibits high intelligence. Algorithms for machine learning (ML) have shown encouraging promise in helping medical practitioners make prompt and precise diagnoses. The supervised machine learning techniques used in this research include principal component analysis (PCA), support vector machines (SVM), random forests (RF), and decision trees (DT).

3. Proposed System

With the following information, we have examined a new dataset on cardiac illnesses.

Table 1. The properties of Heart Disease dataset

Dataset	Instances	Attributes	Missing Values	Yes	No
Heart Disease	1191	11	No	630	561

Some of the models given below will be used in the analysis of heart disorders.

1. Logistic Regression
2. Support Vector Classifier
3. Decision Tree Classifier
4. K-Nearest Neighbors (KNN)
5. Random Forest Classifier
6. Gradient Boosting Classifier

1. Logistic Regression: For classification problems, the supervised learning method known as logistic regression is employed. It is a classification algorithm, not a regression algorithm, as its name suggests. It uses input information to predict discrete class labels (e.g., Yes/No, 0/1, Spam/Not Spam).

2. Support Vector Classifier: For classification tasks, a supervised learning approach called a Support Vector Classifier (SVC) is employed. It operates by determining the best decision border across classes and is based on Support Vector Machines (SVMs).
3. Decision Tree Classifier: For classification problems, a supervised learning approach called a Decision Tree Classifier is employed. It creates a tree-like structure by iteratively dividing the data into subgroups according to feature requirements.
4. K-Nearest Neighbors (KNN): A supervised learning approach for regression and classification is called K-Nearest Neighbors (KNN). A data point's class is determined by the KNN classifier using the majority vote of its closest neighbors.
5. Random Forest Classifier: The supervised ensemble learning algorithm Random Forest is applicable to both regression and classification problems. To increase accuracy and decrease overfitting, it builds several decision trees and combines their outputs.
6. Gradient Boosting Classifier: A potent ensemble learning technique for classification and regression applications is gradient boosting. It creates models one after the other, fixing the mistakes of the earlier models by concentrating on situations that are challenging to categorize.

Table 2. The hyper parameters of machine learning models

Algorithm	Key Hyper parameters
1. Logistic Regression	Penalty = elasticnet
	C = smaller → stronger regularization
	max_iter = 10
2. Support Vector Classifier	C= controls margin softness
	Kernel = poly
	Degree = polynomial
	Gamma = poly
3. Decision Tree Classifier	Criterion = gini
	max_depth = tree depth
	min_samples_split = split a node
	max_features = all
4. K-Nearest Neighbors (KNN)	n_neighbors = 5
	Weights = uniform
	Algorithm = kd_tree
	P = Euclidean
5. Random Forest Classifier	Criterion = entropy
	Bootstrap = yes
6. Gradient Boosting Classifier	n_estimators = 5
	learning_rate = optimal
	Criterion = squared_error
	max_features = 10

4. Methodology

The following list of system requirements includes the operating system, system configuration with processor speed and RAM information, and any graphics hardware or software needed to operate our project. The suggested method is implemented in Python using Jupiter Work Bench, which runs Windows 11 on an AMD Ryzen 5 5500U system unit with Radeon Graphics running at 2.10 GHz and 16.0 GB of RAM.

5. Implementation

First, import the necessary libraries for this project, such as pandas and numpy. We used the CSV dataset, which contains both numerical and nominal values. The missing values are converted using mean value computed from other existing values.

5.1. Pre Processing Data

Understanding the datasets shape: The number of rows and columns in the dataset is provided by the shape attribute. 1190 rows (data points/observations) are shown here. 12 columns (variables/features).
Listing the column Headers: A list of every column name in the Data Frame is provided by the columns attribute.
Data Types of Each Column: Each column's data type is displayed using the data types attribute. The type of data recorded in each column is indicated by these types. data.head(): By default, this command shows the dataset's (data) first five rows. data.tail(): This function shows the dataset's final five rows. It assists in confirming the dataset's completeness and integrity by displaying comparable columns but data from the end of the dataset.

data.isnull().any(): This looks for missing values in every dataset column. If any values are missing, it returns True; otherwise, it returns False. In this instance, it seems that there are no missing data in any of the columns (age, sex, etc.).

data_dup = data.duplicated().any():data.duplicated()checks for duplicate rows in the dataset and produces a boolean series that indicates whether each row is a duplicate (False otherwise, True if it is)..any() determines whether the dataset contains any duplicates, returning False otherwise and True if at least one duplicate occurs. Since data_dup is set to True in this case, duplicate rows are present in the dataset.

data = data.drop_duplicates(): By doing this, all duplicate rows are eliminated from the dataset, leaving only the first instance of each duplicate row. The duplicate-free updated dataset is moved to the variable data.

data_dup = data.duplicated().any() (after dropping duplicates): The algorithm looks for duplicates in the dataset once duplicates have been eliminated. To ensure that there are no duplicates left in the dataset, data_dup is now set to False.

Importing Libraries

importmatplotlib.pyplot as plt: Imports Matplotlib for plotting graphs.

%matplotlib inline:makes certain that plots are shown inline, or inside the Jupyter Notebook, as opposed to in a different window.

importseaborn as sns: Seaborn, a Python package for sophisticated statistical visualizations based on Matplotlib. This snippet does not use it, even if it is imported here. generates unique histograms for every dataset's (data) numerical columns. The distribution of values for each column is shown by the histograms.

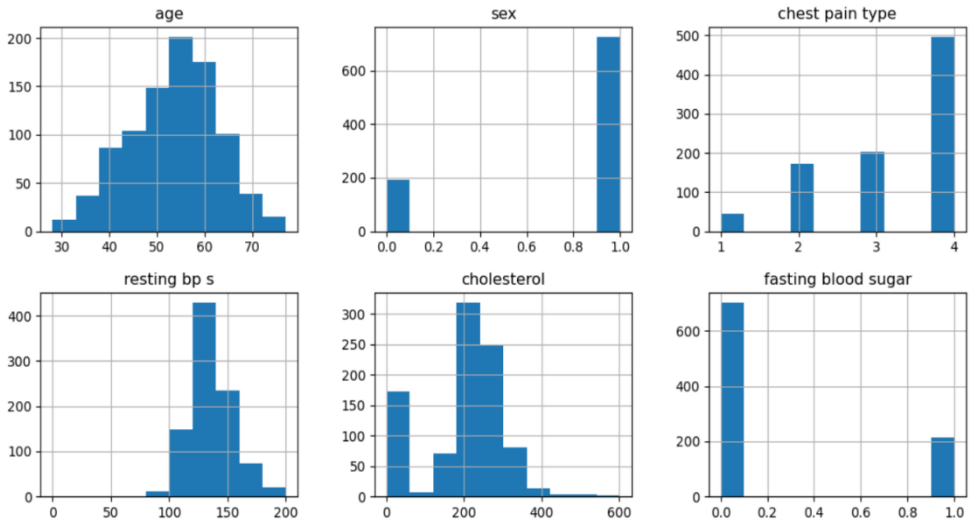


Fig. 1. Individual histograms for dataset columns

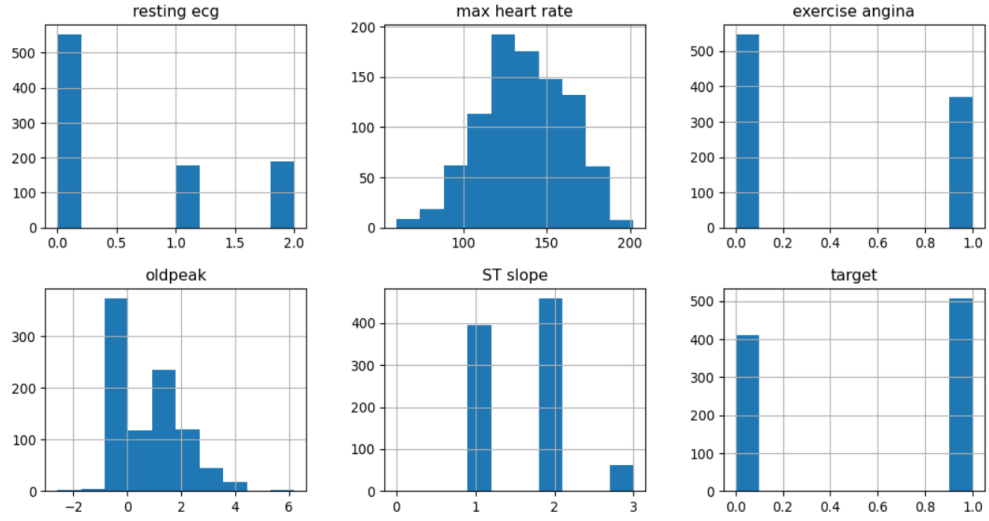


Fig. 2. Individual histograms for dataset columns

5.2. Visualization to check if the dataset is balanced or not

The distribution of the target column is displayed in this visualization using a Seaborn count plot, where 0 and 1 denote two groups (e.g., no heart disease vs. heart disease). In this case, both classes seem to be distributed about evenly, which aids in determining whether the dataset is balanced.

```
[22]: # Visualization to check if the dataset is balanced or not
g = sns.countplot(x='target', data=data)
plt.xlabel('Target')
plt.ylabel('Count')
plt.show()
```

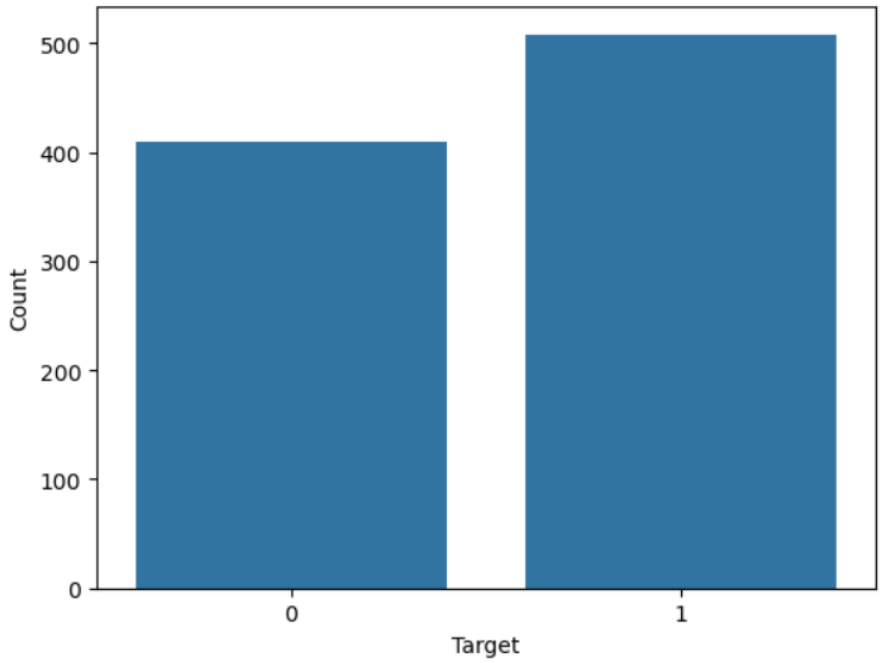


Fig. 3. Visulization of target clasess for class balance nature

5.3. Selecting correlated features using Heatmap

To find features that are highly connected the correlation matrix for each feature in the dataset is calculated and displayed using a heatmap. The heatmap highlights associations using the "RdYlGn" color map and has annotations.

```
[23]: # Selecting correlated features using Heatmap

# Get correlation of all the features of the dataset
corr_matrix = data.corr()
top_corr_features = corr_matrix.index

# Plotting the heatmap
plt.figure(figsize=(20,20))
sns.heatmap(data=data[top_corr_features].corr(), annot=True, cmap='RdYlGn')
plt.show()
```

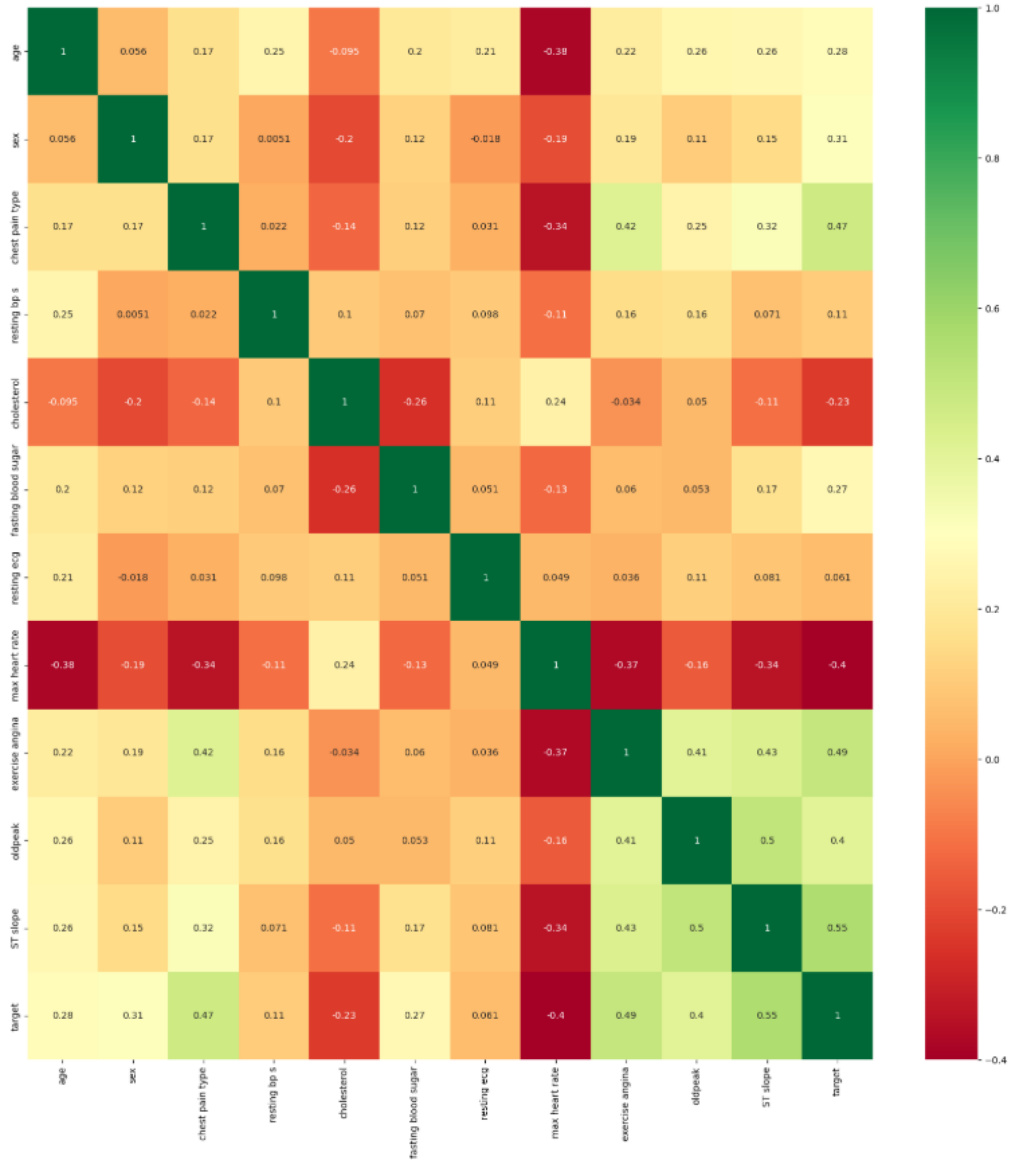


Fig. 4. Heatmap visulization of dataset features.

5.4. Data Processing

Based on the quantity of distinct values, divides dataset columns into continuous (cont_val) and categorical (cate_val) categories. While the remaining columns are continuous, those with 10 or fewer unique values are categorized as categorical.

- **Encoding Categorical Data**

Excludes the sex and target columns when encoding categorical data into dummy/one-hot encoding. By eliminating the first dummy variable for every category, the drop_first=True parameter prevents multicollinearity.

- **Feature Scaling**

After normalizing the continuous variables (cont_val), StandardScaler() sets their mean to 0 and their standard deviation to 1. The original values in data[cont_val] are replaced by the modified data. The top five rows of the updated dataset containing scaled numerical values, prepared for additional processing or modeling, are displayed by the data.head() command.

- **Splitting The dataset into The Training Set And Test Set**

Divides the dataset into test and training sets in order to evaluate and train the model: 1. The target column (y = data['target']) is isolated from the features (x). 2. The data is split into training (x_train, y_train) and testing (x_test, y_test) subsets using the train_test_split function from sklearn, with 20% going to testing. 3. Consistent splitting across

several runs is guaranteed with random_state=42. 4. Scaled features are found in x_train, and the appropriate target labels are found in y_train.

- **Evaluation matrix in machine learning(F1 score)**

Now We used six machine learning models to predict heart disease, most likely involving techniques such as Random Forest, K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting Classifier. Your preprocessed dataset was used to train and test each model, and the F1-score was used to compare how well each performed.

Due to its ensemble nature and resistance to over fitting, the Random Forest Classifier had the strongest capacity to handle the dataset after analyzing these models, as evidenced by its best F1-score. The Random Forest model was then refined for our predictions, utilizing its versatility, resilience, and feature value assessment for binary classification tasks such as heart disease prediction. This guarantees the most dependable results for the use case.

Table 3. Evaluation metrics of machine learning algorithms on F1 Score

Algorithm	**F1 Score**
Logistic Regression	0.862745
Support Vector Machine	0.866667
K-Nearest Neighbors	0.868293
Decision Tree Classifier	0.847291
Random Forest Classifier	**0.895238**
Gradient Boosting Classifier	0.884615

- **Evaluation matrix in machine learning(roc_auc_score)**

Now We used six machine learning models to predict heart disease, most likely involving techniques such as Random Forest, K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting Classifier. Metricsroc_auc_score was used to compare the performance of each model once it was trained and evaluated on your preprocessed dataset.

With the greatest roc_auc_score after examining these models, the Gradient Boosting Classifier stood out as having a strong ability to handle the dataset. This is probably because it is an ensemble model and resists over fitting. After that, we refined the Gradient Boosting Classifier for our predictions, utilizing its versatility, robustness, and feature value assessment for binary classification tasks such as heart disease prediction. This guarantees the most dependable results for the use case.

Table 4. Evaluation metrics of machine learning algorithms on ROC_AUC

Algorithm	**ROC_AUC**
Logistic Regression	0.852773
Support Vector Machine	0.847312
K-Nearest Neighbors	0.857446
Decision Tree Classifier	0.830441
Random Forest Classifier	0.878990
Gradient Boosting Classifier	0.871465

- **Evaluation matrix in machine learning (log_loss)**

Now We used six machine learning models to predict heart disease, most likely involving techniques such as Random Forest, K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting Classifier. Your preprocessed dataset was used to train and test each model, and log_loss metrics were used to compare how well each performed.

Due to its ensemble nature and resistance to over fitting, the Random Forest Classifier had the strongest capacity to handle the dataset after analyzing these models, as evidenced by its highest log_loss. The Random Forest model was then refined for our predictions, utilizing its versatility, resilience, and feature value assessment for binary classification tasks such as heart disease prediction. This guarantees the most dependable results for the use case.

Table 5. Evaluation metrics of machine learning algorithms on Log_loss

Algorithm	**Log_loss**
Logistic Regression	5.484904
Support Vector Machine	5.484904
K-Nearest Neighbors	5.289014
Decision Tree Classifier	6.268461
Random Forest Classifier	4.309567
Gradient Boosting Classifier	4.701346

- **Evaluation matrix in machine learning (R2_Score)**

Now We used six machine learning models to predict heart disease, most likely involving techniques such as Random Forest, K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting Classifier. R2_Score measures were used to compare the performance of each model after it was trained and evaluated on your preprocessed dataset. With the highest R2_Score after examining these models, the Random Forest Classifier stood out as having a great hold on the dataset, probably because it is ensemble in nature and resistant to over fitting. The Random Forest model was then refined for our predictions, utilizing its versatility, resilience, and feature value assessment for binary classification tasks such as heart disease prediction. This guarantees the most dependable results for the use case.

Table 6. Evaluation metrics of machine learning algorithms on R2 Score

Algorithm	**R2 Score**
Logistic Regression	0.374681
Support Vector Machine	0.374681
K-Nearest Neighbors	0.397014
Decision Tree Classifier	0.285350
Random Forest Classifier	0.620342
Gradient Boosting Classifier	0.464013

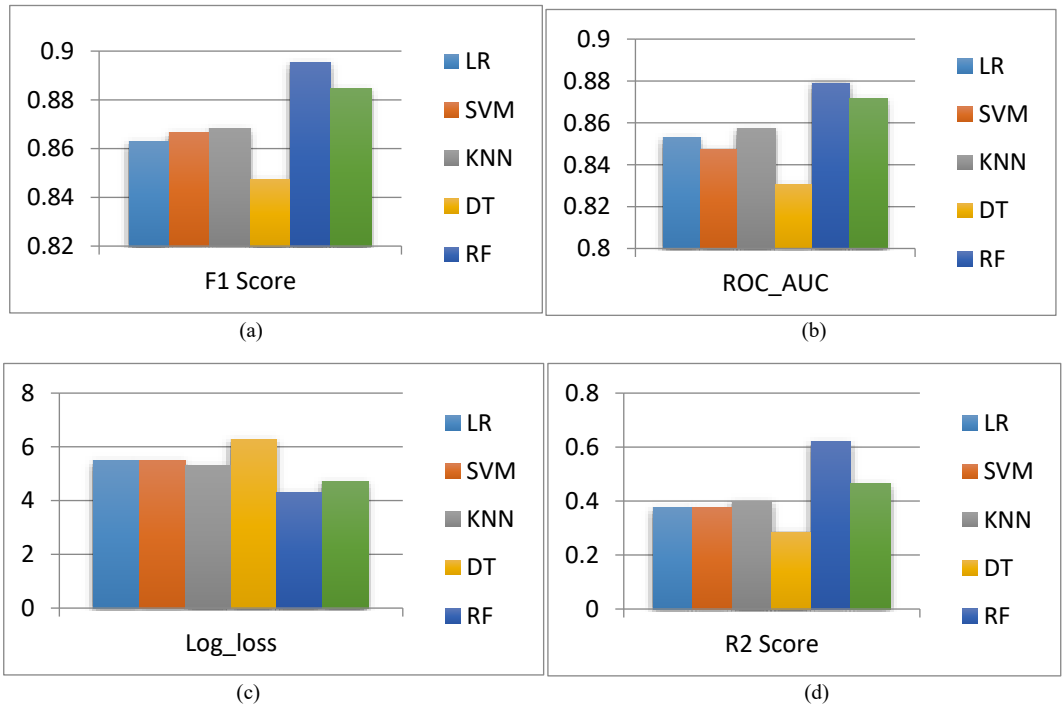


Fig. 5 Bar chart Persentaion for different algorithms (a) F1 score (b) ROC_AUC (c) Log_loss (d) R2 score

Table 7. Significance of the results of the compared algorithms

Algorithm	Log_loss	R^2 Score	Rank
Random Forest Classifier	4.3096	0.6203	1
Gradient Boosting Classifier	4.7013	0.4640	2
K-Nearest Neighbors (KNN)	5.2890	0.3970	3
Logistic Regression	5.4849	0.3747	4
Support Vector Classifier	5.4849	0.3747	4 (tie)
Decision Tree Classifier	6.2685	0.2854	6

The Figure 5 presents the bar charts of F1 score, ROC_AUC, Log_loss and R2 Score. From the above charts one can observe that random forest have performed better in all the metrics. The reason for better performance is due to the group of trees constructed from the dataset and a better and accurate model is used for analysis of the heart disease.

6. Conclusion

The number of heart disease patients has significantly increased in recent years, and heart illness is linked to a high death rate. Furthermore, as technology advanced, several sophisticated devices were created to assist patients in measuring their health at home and estimating their risk of developing heart disease. Using six machine learning models, the study seeks to determine how accurate self-measured physical health indicators are at predicting heart disease when compared to all indicators assessed by medical professionals. Logistics Regression, K Nearest Neighbors, Support Vector Model, Decision Tree, Random Forest, and Gradient Boosting Classifier were among the six models employed to forecast heart disease.

Twelve different test findings and 1189 patients' heart disease risks are included in the database utilized for the study. While the results contained six outcomes that could be tested at home, the all matrices contained all twelve test results. The accuracy score and false negative rate were calculated for each of the five models that were built for the home matrices and all matrices. The findings demonstrated that in all five models, all matrices had greater accuracy scores than home matrices. For five machine learning models, all matrices had false negative rates that were either lower or equivalent to that of results. The results showed that all physical indicators were more accurate in predicting patients' risk of heart disease than home-measured physical health indicators. Therefore, all physical health indicators are preferable for assessing the risk for cardiac illnesses in the absence of future development of home-testable indicators.In future work; we want to extend our proposal in different conceptual directions.

Acknowledgment

The authors would like to thank Krishna University Honorable Vice Chancellor Prof. Ramji Koona for his endless support. We also like to extend our thanks to Krishna University Start-up and Innovation Cell for their support in providing computing and other resources.

References

[1] A. M. Mahmood and M. R. Kuppa, "Early Detection of Clinical Parameters in Heart Disease by Improved Decision Tree Algorithm," *2010 Second Vaagdevi International Conference on Information Technology for Real World Problems*, Warangal, India, 2010, pp. 24-29, doi: 10.1109/VCON.2010.12.

[2] *Huating Sun, Jianan Pan* "Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators", *Journal of Data Analysis and Information Processing*, *Vol.11 No.1, 2023.*

[3] Alkayyali, Z. K., S. Anuar Bin Idris, and Samy S. Abu-Naser. "A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis." *Journal of Theoretical and Applied Information Technology* 101.4 (2023): 1353-1365.

[4] Haq, Amin Ul, et al. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." *Mobile information systems* 2018.1 (2018): 3860146.

[5] A. M. Mahmood and M. R. Kuppa, "Improving the efficiency of biomarker identification using expert knowledge," *Trendz in Information Sciences &Computing(TISC2010)*, Chennai, India, 2010, pp. 111-115, doi: 10.1109/TISC.2010.5714618.

[6] Mahmood, Ali Mirza, and MrithyumjayaRaoKuppa. "Increasing generalization accuracy by using multivariate statistical method." *International Journal of Computer Systems Science & Engineering* 26.2 (2011): 97-102.

[7] A. M. Mahmood and M. R. Kuppa, "Improving the efficiency of biomarker identification using expert knowledge," *Trendz in Information Sciences &Computing(TISC2010)*, Chennai, India, 2010, pp. 111-115, doi: 10.1109/TISC.2010.5714618.

[8] G. Dinesh, Dr Ali MirzaMahmood , “A Novel Corona Virus Detection and Validation Measures using Machine Learning Techniques”A Advancement of Computer Technology and its Applications, Volume 5 Issue 1, 10.5281/zenodo.6476714

[9] Dr. Mahmood Ali Mirza,&Faheem Ali Mirza. (2025). Classification Approach for Analysis of Weather Dataset with Different Training Strategies. Recent Trends in Information Technology and Its Application, 8(2), 4–13.https://doi.org/10.5281/zenodo.14997848.
[10] Thota, N. R., and Vasumathi, D. (2023). A Distinctive Ensemble Deep Learning Model for Brain Tumor MRI Image Classification. *i-manager's Journal on Artificial Intelligence & Machine Learning*, 1(2), 12-21. https://doi.org/10.26634/jaim.1.2.19281
[11] NR Thota, D Vasumathi , "Wasserstein GAN-gradient penalty with deep transfer learning based Alzheimer disease classification on 3D MRI scans", i-manager's Journal on Image Processing, Vol 9, Issue 4,, 2022
[12] NarasimhaRaoThota, D Vasumathi," BRAIN TUMOR CLASSIFICATION WITH CONVLSTM USING NAKAGAMI PARAMETRIC IMAGING AND BAYESIAN FUZZY CLUSTERING", International Journal of Engineering Applied Sciences and Technology, Vol 8, Issue 10,, 2024.
[13] Kumar, N. S., Rao, K. N., Govardhan, A., Reddy, K. S., &Mahmood, A. M. (2014). Undersampled K-means approach for handling imbalanced distributed data. *Progress in Artificial Intelligence*, *3*(1), 29-38.
[14] Mahmood, A. M. A Classification Approach for Identification of Glass using C4. 5 Algorithm. Advancement of Computer Technology and its Applications Volume 7 Issue 3 e-ISSN: 2584-1262 DOI: https://doi.org/10.5281/zenodo.11401359.
[15] Mahmood, A. M. (2015). Class imbalance learning in data mining–a survey. *International Journal of Communication Technology for Social Networking Services*, *3*(2), 17-38.
[16] Mahmood, A. M., Satuluri, N., &Kuppa, M. R. (2011). An Overview of recent and traditional decision tree classifiers in machine learning. *International Journal of Research and Reviews in Ad Hoc Networks*, *1*(1), 2011.
[17] Mahmood, A. M., &Kuppa, M. R. (2010, December). A novel pruning approach using expert knowledge. In *INTERACT-2010* (pp. 33-38). IEEE.
[18] Mirza, M. A., Babu, G. S., Ajay, G. P., Rohit, D., &Anand, C. BE-Abhaya: A Next Gen safety Application for Emergency Response and Risk Mitigation. Advancement of Computer Technology and its Applications Volume 8 Issue 2 e-ISSN: 2584-1262 DOI: https://doi.org/10.5281/zenodo.14970112
[19] Mahmood, A. M., Kowshik, B., &Mirza, F. A. Artificial Intelligence Techniques, Methods and Framework: An Epigrammatic Analysis. Advancement of Computer Technology and its Applications Volume 8 Issue 1 e-ISSN: 2584-1262 DOI: https://doi.org/10.5281/zenodo.13999765
[20] Tyagi, N., Jain, P. (2024). Heart Disease Prediction Using Machine Learning Algorithms. In: Tiwari, S., Trivedi, M.C., Kolhe, M.L., Singh, B.K. (eds) Advances in Data and Information Sciences. ICDIS 2024. Lecture Notes in Networks and Systems, vol 1127. Springer, Singapore.https://doi.org/10.1007/978-981-97-7360-2_5
[21] Mishra, A. *et al.* (2024). Heart Disease Prediction by Machine Learning. In: Mandal, J.K., De, D. (eds) Machine Learning for Social Transformation. EAIT 2024. Lecture Notes in Networks and Systems, vol 1131. Springer, Singapore. https://doi.org/10.1007/978-981-97-7532-3_25
[22] Ekle, F.A., Shidali, V., Ochogwu, R.E. *et al.* Machine Learning models for heart disease prediction and dietary lifestyle change therapy recommendation: a systematic review. *Discov Artif Intell* 4, 113 (2024). https://doi.org/10.1007/s44163-024-00181-w
[23] KuldeepVayadande, ArnavDhiwar, Darpan Khadke, Rohan Golawar, Sarwesh Khairnar, Sarthak Wakchoure, Sumeet Bhoite. Published in: Techno-societal 2022, Publisher: Springer International Publishing
[24] Mallikarjunamallu, K., Syed, K. (2024). A Review on Heart Diseases Using Machine Learning and Deep Learning Techniques. In: Pant, M., Deep, K., Nagar, A. (eds) Proceedings of the 12th International Conference on Soft Computing for Problem Solving. SocProS 2023. Lecture Notes in Networks and Systems, vol 995. Springer, Singapore. https://doi.org/10.1007/978-981-97-3292-0_45
[25] G. Sakthipriya, Y. Suresh, C. Varnisha, R. Sindhu, R. Shivraj, Published in: Innovations in VLSI, Signal Processing and Computational Technologies, Publisher: Springer Nature Singapore
[26] Bhavekar, Girish&Goswami, Agam & Chafle, Pratiksha & Gaikwad, Amit & Vyawahare, Harsha. (2024). Heart disease prediction using machine learning, deep Learning and optimization techniques-A semantic review. Multimedia Tools and Applications. 83. 86895-86922. 10.1007/s11042-024-19680-0.
[27] Amrit Singh, HarisankarMahapatra, Anil Kumar Biswal, Madhumita Mahapatra, Debabrata Singh, Milan Samantaray, Heart Disease Detection Using Machine Learning Models, Procedia Computer Science, Volume 235, 2024, Pages 937-947, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2024.04.089.
[28] Breiman, Leo. "Bagging predictors." *Machine learning* 24 (1996): 123-140.
[29] Parhami, Behrooz. "Voting algorithms." *IEEE transactions on reliability* 43.4 (1994): 617-629.
[30] Schapire, Robert E. "A brief introduction to boosting." *Ijcai*. Vol. 99. No. 999. 1999.
[31] Rigatti, Steven J. "Random forest." *Journal of Insurance Medicine* 47.1 (2017): 31-39.

Authors' Profiles



Dr. Ali Mirza Mahmood received his PhD in 2018 in Computer Science and Engineering at Acharya Nagarjuna University. He received his Master's Degree in Computer Science in 2003. Now, he is an Assistant Professor in Krishna University College of Engineering and Technology, Machilipatnam (India). His current research interest includes Data Mining and Knowledge Discovery, Machine Learning, and Artificial Intelligence. He is on Editorial board and reviewer for many reputed journals such as Springer, Elsevier, IEEE etc. He has published more than 40 papers in reputed journals and conferences.



Jaddu Lohith Kumaris pursuing his final year Graduation in Computer Science and Engineering from Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh. He has done many online certifications from edx International platform.



Gandavarapu Mohan Sai is a Graduate in Computer Science and Engineering from Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh. He has done many online certifications from edx International platform.



Kongora Venkat Chowdary is a Graduate in Computer Science and Engineering from Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh. He has done many online certifications from edx International platform.



Karnaati Kranthi Kiran is a Graduate in Computer Science and Engineering from Krishna University College of Engineering and Technology, Machilipatnam, AndhraPradesh. He has done many online certifications from edx International platform.



Faheem Ali Mirza is pursuing his Graduation in Engineering in Information Technology from Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh. He has authored four paper in international journals. He has done many online certifications from edx International platform.

How to cite this paper: Mahmood Ali Mirza, Jaddu Lohith Kumar, Gandavarapu Mohan Sai, Kongora Venkat Chowdary, Karnaati Kranthi Kiran, Faheem Ali Mirza, "A Systematic Approach for Heart Disease Analysis using Machine Learning Algorithms", International Journal of Engineering and Manufacturing (IJEM), Vol.16, No.2, pp.214-225, 2026. DOI:10.5815/ijem.2026.02.14