IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, April 2025

Impact Factor: 7.67

Comparative Analysis of Machine Learning Models for Heart Disease Prediction

Ms. Sanjivani M. Bhade[1], Dr. A. P. Thakare[2], Dr. Vaishali A.Thakare[3]

M.E. Student, SIPNA C.O.E.T. Amravati, Maharashtra, India[1]

Professor, SIPNA C.O.E.T. Amravati, Maharashtra, India[2]

Lecturer, Dental College and Hospital Amravati, Maharashtra, India[3]

Abstract: *The diagnosis and prognosis of cardiovascular disease play a vital role in ensuring accurate classification, which assists cardiologists in providing appropriate treatment to patients. The adoption of machine learning in the medical field has grown significantly due to its ability to detect patterns from data. Leveraging machine learning for cardiovascular disease classification can help reduce the risk of misdiagnosis. This study presents a model designed to accurately predict cardiovascular disease occurrences, ultimately minimizing fatalities. The proposed method utilizes k-modes clustering with Huang initialization to enhance classification accuracy. Machine learning algorithms such as Random Forest (RF), Decision Tree Classifier (DT), Multilayer Perceptron (MP), and XGBoost (XGB) are employed. Hyperparameter tuning using GridSearchCV was conducted to optimize model performance. The model was applied to a real-world Kaggle dataset comprising 70,000 instances, with an 80:20 train-test split.*

Keywords: Heart disease; machine learning; k-modes clustering; classification; multilayer perceptron; model assessment

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of morbidity and mortality worldwide, contributing to over 70% of all deaths. The 2017 Global Burden of Disease study reports that CVDs account for approximately 43% of total fatalities. In high-income countries, common risk factors for heart disease include poor dietary habits, smoking, excessive sugar intake, and obesity. However, the prevalence of chronic illnesses is also increasing in low- and middle-income countries. Between 2010 and 2015, the global economic impact of cardiovascular diseases was estimated to be around USD 3.7 trillion.

Furthermore, diagnostic technologies such as electrocardiograms and CT scans, which are crucial for detecting coronary heart disease, can often be too expensive and impractical for many individuals. This limitation alone has contributed to the deaths of 17 million people. Additionally, cardiovascular disease accounts for 25% to 30% of companies' annual medical expenses. Therefore, early detection of heart disease is essential to reducing its impact on society. Utilizing data mining and machine learning techniques can help predict the likelihood of heart disease, mitigating both its physical and financial burden on individuals and institutions. According to WHO projections, the total number of deaths from CVDs is expected to rise to 23.6 million by 2030, with heart disease and stroke remaining the leading causes. Implementing effective predictive measures is crucial to saving lives and reducing healthcare costs.

Cardiovascular disease (CVD) remains a leading cause of morbidity and mortality worldwide, contributing to over 70% of global deaths. The 2017 Global Burden of Disease Study indicates that CVD accounts for more than 43% of all fatalities. Key risk factors for heart disease include poor diet, tobacco use, excessive sugar consumption, and obesity, which are prevalent in high-income countries. However, the incidence of chronic diseases is also rising in low- and middle-income nations. The global economic burden of CVDs was estimated at approximately USD 3.7 trillion between 2010 and 2015.

Moreover, diagnostic tools such as electrocardiograms and CT scans, which are essential for detecting coronary heart disease, are often prohibitively expensive and inaccessible in many low- and middle-income countries. Early detection is therefore critical to reducing the physical and financial impact of CVDs on individuals and organizations. According

DOI: 10.48175/IJARSCT-24974

to a WHO report, CVD-related deaths are projected to reach 23.6 million by 2030, with heart disease and stroke being the primary causes. To mitigate this growing crisis, the application of data mining and machine learning techniques is essential for predicting heart disease risk, ultimately helping to save lives and reduce healthcare costs worldwide.

In the medical field, an enormous amount of data is generated daily. By utilizing data mining techniques, hidden patterns within this data can be uncovered and leveraged for clinical diagnosis. Over the past few decades, data mining has proven to be an essential tool in healthcare. Various factors, including diabetes, high blood pressure, high cholesterol, and abnormal pulse rate, must be considered when predicting heart disease. However, incomplete medical data often pose a challenge, affecting the accuracy of heart disease predictions.

Machine learning has become increasingly important in healthcare, enabling the diagnosis, detection, and prediction of various diseases. In recent years, there has been a growing focus on applying data mining and machine learning techniques to assess the likelihood of disease development. While existing studies have explored the use of these techniques for disease prediction, many have struggled to achieve highly accurate results. This study aims to enhance heart disease prediction by developing a more precise and reliable model for assessing the risk of heart disease in individuals.

This study aims to evaluate the effectiveness of various machine learning algorithms in predicting heart disease. To achieve this, we utilized multiple techniques, including Random Forest, Decision Tree Classifier, Multilayer Perceptron, and XGBoost, to develop predictive models. To enhance model convergence, we applied k-modes clustering for dataset preprocessing and scaling. The dataset used in this research is publicly available on Kaggle, and all computations, preprocessing, and visualizations were performed on Google Colab using Python.

Previous studies have reported accuracy rates of up to 94% in heart disease prediction using machine learning. However, these studies often relied on small sample sizes, limiting the generalizability of their findings. Our research seeks to overcome this limitation by employing a larger and more diverse dataset, thereby improving the robustness and applicability of the results to a broader population.

II. LITERATURE SURVEY

In recent years, the healthcare industry has experienced significant advancements in data mining and machine learning. These technologies have been widely adopted and have proven effective in various medical applications, particularly in cardiology. The rapid growth of medical data has provided researchers with an opportunity to develop and refine new algorithms in this field. Heart disease remains a leading cause of mortality in developing countries, making the identification of risk factors and early symptoms a critical area of research. The application of data mining and machine learning techniques has the potential to enhance early detection and prevention efforts for heart disease.

Narain et al. (2016) developed a machine learning-based system using a quantum neural network to improve the accuracy of cardiovascular disease (CVD) prediction compared to the Framingham Risk Score (FRS). Using data from 689 individuals and a validation dataset from the Framingham study, the system achieved an impressive 98.57% accuracy, far surpassing the FRS accuracy of 19.22%. The study suggests that this approach could help physicians with risk assessment, early diagnosis, and improved treatment strategies.

Shah et al. (2020) developed a machine learning model to predict cardiovascular disease using the Cleveland Heart Disease dataset, which included 303 instances and 17 attributes. They applied several classification techniques, including Naïve Bayes, Decision Tree, Random Forest, and k-Nearest Neighbors (KNN). The KNN model achieved the highest accuracy at 90.8%, emphasizing the potential of machine learning in disease prediction and the importance of selecting optimal models for better performance.

A study by Drod et al. (2022) used machine learning (ML) techniques to identify key risk factors for cardiovascular disease (CVD) in 191 patients with metabolic-associated fatty liver disease (MAFLD). Using methods like logistic regression, feature ranking, and principal component analysis (PCA), the study found that hypercholesterolemia, plaque scores, and diabetes duration were the most critical factors. The ML model accurately classified 85.11% of high-risk and 79.17% of low-risk patients, with an AUC of 0.87, highlighting the potential of ML in detecting CVD risk in MAFLD patients using basic clinical parameters.

DOI: 10.48175/IJARSCT-24974

Alotalibi (2019) investigated the effectiveness of machine learning (ML) techniques in predicting heart failure using data from the Cleveland Clinic Foundation. Various ML algorithms, including Decision Tree, Logistic Regression, Random Forest, Naïve Bayes, and Support Vector Machine (SVM), were applied with a 10-fold cross-validation method. The Decision Tree algorithm achieved the highest accuracy at 93.19%, followed closely by SVM at 92.30%. The study highlights the potential of ML techniques in heart failure prediction, with the Decision Tree algorithm showing the most promise for future research.

Hasan and Bao (2020) compared three feature selection techniques—filter, wrapper, and embedding—to predict cardiovascular disease. They evaluated various models, including Random Forest, SVC, KNN, Naïve Bayes, and XGBoost, using an ANN as a benchmark. The study found that XGBoost combined with the wrapper method achieved the highest accuracy (73.74%), followed by SVC (73.18%) and ANN (73.20%).

One major limitation of previous studies is their reliance on small datasets, increasing the risk of overfitting and limiting model generalizability. In contrast, our research utilizes a cardiovascular disease dataset comprising 70,000 patients and 11 features, significantly reducing the likelihood of overfitting. It provides a summary of cardiovascular disease prediction studies conducted on large datasets, further demonstrating the advantages of using a comprehensive dataset for enhanced predictive performance.

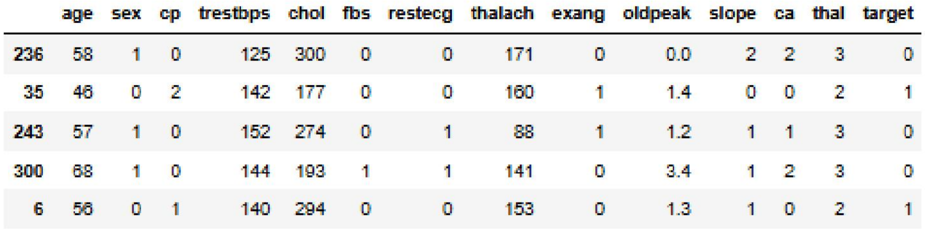
III. METHODOLOGY

This study aims to predict the likelihood of heart disease using a computerized heart disease prediction system, benefiting both medical professionals and patients. To achieve this, we applied various machine learning algorithms to a dataset and analyzed the results.

To refine the methodology, we plan to preprocess the data by cleaning it, removing irrelevant information, and incorporating additional features such as Mean Arterial Pressure (MAP) and Body Mass Index (BMI). The dataset will then be stratified based on gender, followed by the application of k-modes clustering. Finally, the model will be trained using the processed data. These enhancements are expected to improve accuracy and overall model performance.

3.1. Data Source

The dataset used in this study, as referenced in [23], consists of 10,000 patient records with 14 unique features, as outlined in fig. 3.1 These features include age, gender, systolic blood pressure, and diastolic blood pressure. The target variable, "cardio," signifies whether a patient has cardiovascular disease (denoted as 1) or is healthy (denoted as 0).



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
236	58	1	0	125	300	0	0	171	0	0.0	2	2	3	0
35	46	0	2	142	177	0	0	160	1	1.4	0	0	2	1
243	57	1	0	152	274	0	1	88	1	1.2	1	1	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1

Fig. 3.1 Sample Dataset with 14 features

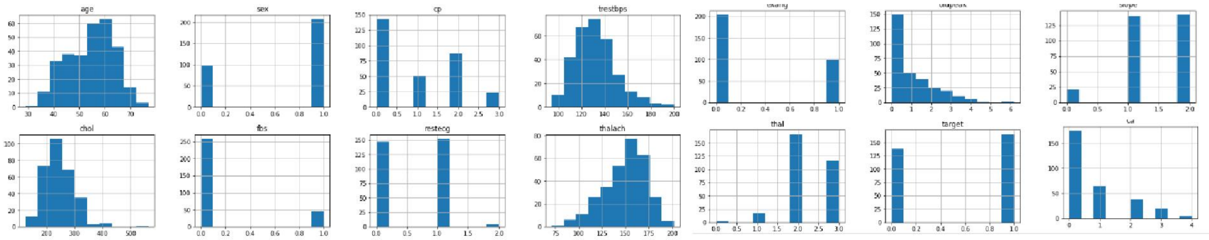


Fig. 3.2 Graphs of dataset

DOI: 10.48175/IJARSCT-24974

3.2. Removing Outliers

As illustrated in Fig.1, the dataset contains noticeable outliers, likely resulting from data entry errors. Eliminating these outliers can enhance the performance of our predictive model. To address this, we manually removed all instances of ap_hi, ap_lo, weight, and height that fell outside the 2.5% to 97.5% range. This data cleaning process reduced the total number of records from 70,000 to 57,155.

3.3. Feature Selection and Reduction

We propose using binning as a technique to convert continuous variables, such as age, into categorical variables to enhance the performance and interpretability of classification algorithms. By grouping continuous data into discrete categories or bins, the algorithm can better differentiate between various data classes based on specific input values. For example, if the input variable is "Age Group" with categories such as "Young," "Middle-aged," and "Elderly," the classification algorithm can utilize these predefined groups to categorize individuals accordingly [24].

Moreover, transforming continuous variables into categorical ones through binning improves result interpretability, making it easier to understand the relationship between input features and output classes. In contrast, working directly with continuous numerical values can be more challenging for classification algorithms, as they may need to establish arbitrary decision boundaries between different categories [25].

In this study, we applied binning to the age attribute in a patient dataset. Initially, age was recorded in days, but for improved analysis and prediction, it was converted to years by dividing by 365. The age data was then grouped into 5-year intervals, ranging from 0–20 to 95–100. Since the dataset includes patients aged between 30 and 65, the 30–35 age group was labeled as 0, while the 60–65 group was labeled as 6.

Additionally, other continuous attributes such as height, weight, ap_hi, and ap_lo were also transformed into categorical variables. The findings of this study indicate that converting continuous data into categorical values through binning enhances both the performance and interpretability of classification algorithms.

3.4. Clustering

Clustering is a machine learning technique used to group similar instances, with k-means being a common algorithm. However, k-means is ineffective for categorical data, which led to the development of the k-modes algorithm by Huang in 1997. K-modes uses dissimilarity measures and cluster modes instead of numerical distances, making it better suited for categorical datasets.

Since the dataset in question has been transformed into categorical data, k-modes clustering will be used. The elbow curve method with Huang initialization will help determine the optimal number of clusters by plotting costs for different cluster numbers and identifying the point where adding more clusters no longer improves the model significantly.

Additionally, splitting the dataset by gender can improve prediction accuracy by accounting for biological differences between men and women that affect disease manifestation and progression. Since men and women exhibit different heart disease symptoms, risk factors, and prevalence rates, analyzing data separately may uncover distinct patterns that could be missed in a combined analysis.

3.5. Correlation Table

Additionally, a correlation table is created to analyze the relationships between different categoriesmean arterial pressure (MAP_Class), cholesterol levels, and age exhibit strong correlations. This matrix also helps identify intra-feature dependencies within the dataset.

3.6. Modeling

The dataset is divided into a training set (80%) and a testing set (20%). The model is trained using the training set, and its performance is evaluated on the testing set. Various classifiers, including the decision tree classifier, random forest classifier, multilayer perceptron, and XGBoost, are applied to the clustered dataset to measure their effectiveness. Each classifier's performance is assessed using metrics such as accuracy, precision, recall, and F-measure.

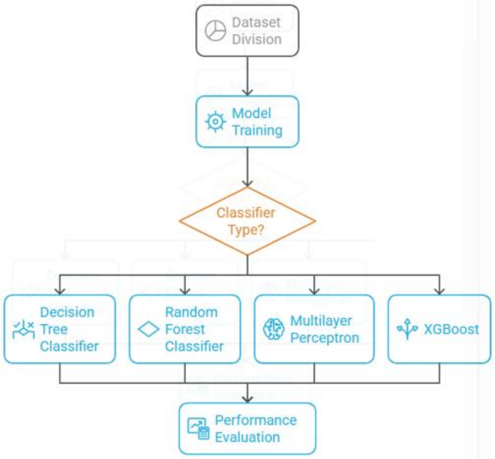


Fig. 3.3 Machine Learning Model Evaluation Process

3.6.1. Decision Tree Classifier

Decision trees are hierarchical structures used to process large datasets efficiently. They are often represented as flowcharts, where inner nodes denote dataset attributes, and outer branches indicate possible outcomes. Decision trees are widely favoured for their simplicity, reliability, and ease of interpretation. The predicted class label originates from the tree's root, with subsequent steps determined by comparing the root attribute's value to the dataset's records. The decision path follows the corresponding branch based on the comparison results. When training examples are divided into smaller subsets, entropy changes, and this change is measured as information gain, which helps determine the best attribute for splitting the data.

3.6.2. Random Forest

The random forest algorithm [13] is a supervised classification technique that combines multiple decision trees to improve prediction accuracy. Each tree within the random forest independently predicts a class, and the final prediction is determined by majority voting. This method overcomes the limitations of individual decision trees by reducing overfitting and enhancing model reliability. Even when applied to large datasets with missing values, the random forest algorithm can still deliver consistent results. Additionally, the samples generated by decision trees within the model can be stored and used with various data types [31]. In the study referenced in [7], the random forest classifier achieved a test accuracy of 73% and a validation accuracy of 72% using 500 estimators, a maximum depth of 4, and a random state of 1.

3.6.3. Multilayer Perceptron

A multilayer perceptron (MLP) is a type of artificial neural network composed of multiple layers, making it more effective at handling nonlinear problems compared to a single-layer perceptron, which is limited to linear classifications. MLP is commonly used for solving complex problems and is an example of a feedforward neural network with multiple layers [32].

Unlike single-layer perceptrons that rely solely on step functions, MLPs utilize other activation functions, such as the sigmoid function, which enables smooth transitions rather than abrupt decision boundaries [33]. The learning process in MLP involves adjusting the perceptron's weights to minimize errors, a task achieved through the backpropagation technique, which helps reduce mean squared error (MSE).

3.6.4. XGBoost

XGBoost [14] is an advanced version of gradient-boosted decision trees, designed to improve prediction accuracy through sequential learning. In this algorithm, decision trees are built in succession, with each tree learning from the

DOI: 10.48175/IJARSCT-24974

errors of the previous one. Independent variables are assigned weights, which influence the tree's predictions. If a tree makes an incorrect prediction, the importance of the relevant variables is increased and carried forward to the next tree. The outputs of all individual trees are then combined to create a stronger and more accurate model.

In a study by [34], the XGBoost model achieved an accuracy of 73% using parameters such as a learning rate of 0.1, a maximum depth of 4, 100 estimators, and 10-fold cross-validation. The model was trained on 49,000 instances and tested on 21,000 instances from a dataset of 70,000 cardiovascular disease cases.

The achieved accuracy rates are as follows: Decision Tree – 86.37% (with cross-validation) and 86.53% (without cross-validation), XGBoost – 86.87% (with cross-validation) and 87.02% (without cross-validation), Random Forest – 87.05% (with cross-validation) and 86.92% (without cross-validation), and Multilayer Perceptron – 87.28% (with cross-validation) and 86.94% (without cross-validation). The AUC (Area Under the Curve) values for the models are: Decision Tree – 0.94, XGBoost – 0.95, Random Forest – 0.95, and Multilayer Perceptron – 0.95. The study concludes that the Multilayer Perceptron model with cross-validation outperformed all other algorithms, achieving the highest accuracy of 87.28%.

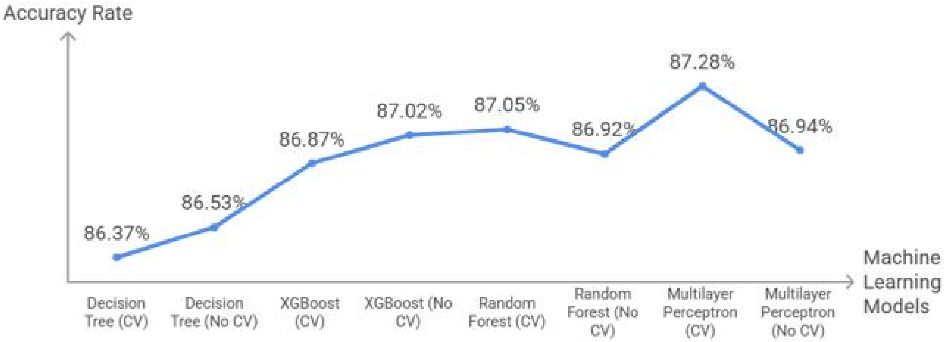


Fig.3.4 Accuracy rates of Machine Learning Models

IV. RESULTS

This study was conducted using Google Colab on a Ryzen 7 4800-H processor with 16 GB of RAM. Initially, the dataset contained 70,000 rows and 12 attributes, but after data cleaning and preprocessing, it was refined to approximately 59,000 rows and 11 attributes. Since all attributes were categorical, outliers were removed to enhance model efficiency. The machine learning algorithms applied in this research included random forest, decision tree, multilayer perceptron (MLP), and XGBoost classifiers.

The performance of these models was evaluated using various metrics, including precision, recall, accuracy, F1 score, and the area under the ROC curve (AUC). The dataset was divided into two subsets: 80% for training and 20% for testing.

For hyperparameter tuning, we utilized an automated approach with the GridSearchCV method. This method, implemented in the scikit-learn library, systematically searches for the optimal hyperparameters by evaluating different combinations using k-fold cross-validation.

Various machine learning classifiers, including MLP, random forest, decision tree, and XGBoost, were applied to predict cardiovascular disease after hyperparameter tuning. The results indicate that the MLP algorithm achieved the highest cross-validation accuracy of 87.28%, along with high recall, precision, F1 score, and AUC values of 84.85, 88.70, 86.71, and 0.95, respectively. All classifiers demonstrated an accuracy exceeding 86.5%. Through hyperparameter tuning with GridSearchCV, the accuracy of the random forest algorithm improved by 0.5%, increasing from 86.48% to 86.90%, while the XGBoost algorithm's accuracy increased by 0.6%, from 86.4% to 87.02%.

V. CONCLUSIONS

The main goal of this study was to classify heart disease using various machine learning models on a real-world dataset. The k-modes clustering algorithm was applied to a dataset of heart disease patients to predict disease presence. Data preprocessing included converting the age attribute from days to years and categorizing it into 5-year intervals. Additionally, diastolic and systolic blood pressure values were divided into bins of 10 intervals. To account for gender-specific differences in heart disease characteristics and progression, the dataset was also split by gender.

The optimal number of clusters for male and female datasets was determined using the elbow curve method. The results showed that the multilayer perceptron (MLP) model achieved the highest accuracy at 87.23%. These findings highlight the effectiveness of k-modes clustering in accurately predicting heart disease, suggesting its potential as a valuable tool for developing targeted diagnostic and treatment strategies.

This study utilized the Kaggle cardiovascular disease dataset, consisting of 70,000 instances, with all machine learning algorithms implemented on Google Colab. The classification models achieved accuracies exceeding 86%, with decision trees recording the lowest accuracy at 86.37%, while the MLP model achieved the highest accuracy, as previously mentioned.

REFERENCES

[1]. Estes, C.; Anstee, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. J. Hepatol. 2018, 69, 896–904.

[2]. Droz˙dz˙ , K.;et al. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240.

[3]. Murthy, H.S.N.; Meenakshi, M. Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease. In Proceedings of the International Conference on Circuits, Communication, Control and Computing, Bangalore, India, 21–22 November 2014; pp. 329–332.

[4]. Benjamin, E.J.; et al. Heart disease and stroke statistics—2019 update: A report from the American heart association. Circulation 2019, 139, e56–e528.

[5]. Shorewala, V. Early detection of coronary heart disease using ensemble techniques. Inform. Med. Unlocked 2021, 26, 100655.

[6]. Mozaffarian, D.; et al. Heart disease and stroke statistics—2015 update: A report from the American Heart Association. Circulation 2015, 131, e29–e322.

[7]. Maiga, J.; Hungilo, G.G.; Pranowo. Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data. In Proceedings of the 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 24–25 October 2019; pp. 45–48.

[8]. Li, J.; Loerbroks, A.; Bosma, H.; Angerer, P.Work stress and cardiovascular disease: A life course perspective. J. Occup. Health 2016, 58, 216–219.

[9]. Purushottam; Saxena, K.; Sharma, R. Efficient Heart Disease Prediction System. Procedia Comput. Sci. 2016, 85, 962–969.

[10]. Soni, J.; Ansari, U.; Sharma, D.; Soni, S. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.Int. J. Comput. Appl. 2011, 17, 43–48.

[11]. Mohan, S.; Thirumalai, C.; Srivastava, G. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. IEEE Access 2019, 7, 81542–81554.

[12]. Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G. Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med. 2020, 7, 1638–1645.

[13]. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.

[14]. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.

[15]. Gietzelt, M.;Wolf, K.-H.; Marschollek, M.; Haux, R. Performance comparison of accelerometer calibration algorithms based on 3D-ellipsoid fitting methods. Comput. Methods Programs Biomed. 2013, 111, 62–71.

[16]. K, V.; Singaraju, J. Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks. Int. J. Comput. Appl. 2011, 19, 6–12.

[17]. Narin, A.; Isler, Y.; Ozer, M. Early prediction of Paroxysmal Atrial Fibrillation using frequency domain measures of heart rate variability. In Proceedings of the 2016Medical Technologies National Congress (TIPTEKNO), Antalya, Turkey, 27–29 October 2016.

[18]. Shah, D.; Patel, S.; Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN Comput. Sci. 2020, 1, 345.

[19]. Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. Int. J. Adv. Comput. Sci. Appl. 2019,10, 261–268.

[20]. Hasan, N.; Bao, Y. Comparing different feature selection algorithms for cardiovascular disease prediction. Health Technol. 2020, 11, 49–62.

[21]. Ouf, S.; ElSeddawy, A.I.B. A proposed paradigm for intelligent heart disease prediction system using data mining techniques. J. Southwest Jiaotong Univ. 2021, 56, 220–240.

[22]. Khan, I.H.; Mondal, M.R.H. Data-Driven Diagnosis of Heart Disease. Int. J. Comput. Appl. 2020, 176, 46–54.

[23]. Kaggle Cardiovascular Disease Dataset. Available online: https://www.kaggle.com/datasets/sulianova/cardiovascular-diseasedataset (accessed on 1 November 2022).

[24]. Han, J.A.; Kamber, M. Data Mining: Concepts and Techniques, 3rd ed.; Morgan Kaufmann Publishers: San Francisco, CA, USA, 2011.

[25]. Rivero, R.; Garcia, P. A Comparative Study of Discretization Techniques for Naive Bayes Classifiers. IEEE Trans. Knowl. Data Eng. 2009, 21, 674–688.

[26]. Khan, S.S.; Ning, H.; Wilkins, J.T.; Allen, N.; Carnethon, M.; Berry, J.D.; Sweis, R.N.; Lloyd-Jones, D.M. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. JAMA Cardiol. 2018, 3, 280–287.

[27]. Kengne, A.-P.; Czernichow, et al. Blood Pressure Variables and Cardiovascular Risk. Hypertension 2009, 54, 399–404.

[28]. Yu, D.; Zhao, Z.; Simmons, D. Interaction between Mean Arterial Pressure and HbA1c in Prediction of Cardiovascular Disease Hospitalisation: A Population-Based Case-Control Study. J. Diabetes Res. 2016, 2016, 8714745.

[29]. Huang, Z. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. DMKD 1997, 3, 34–39.

[30]. Maas, A.H.; Appelman, Y.E. Gender differences in coronary heart disease. Neth. Heart J. 2010, 18, 598–602.

[31]. Bhunia, P.K.; Debnath, A.; Mondal, P.; D E, M.; Ganguly, K.; Rakshit, P. Heart Disease Prediction using Machine Learning. Int. J. Eng. Res. Technol. 2021, 9.

[32]. Mohanty, M.D.; Mohanty, M.N. Verbal sentiment analysis and detection using recurrent neural network. In Advanced Data Mining Tools and Methods for Social Computing; Academic Press: Cambridge, MA, USA, 2022; pp. 85–106.

[33]. Menzies, T.; Kocagüneli, E.; Minku, L.; Peters, F.; Turhan, B. Using goals in model-based reasoning. In Sharing Data and Models in Software Engineering; Morgan Kaufmann: San Francisco, CA, USA, 2015; pp. 321–353.

[34]. Fayez, M.; Kurnaz, S. Novel method for diagnosis diseases using advanced high-performance machine learning system. Appl. Nanosci. 2021.

[35]. Hassan, C.A.U.; Iqbal, J.; Irfan, R.; Hussain, S.; Algarni, A.D.; Bukhari, S.S.H.; Alturki, N.; Ullah, S.S. Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. Sensors 2022, 22, 7227.

[36]. Subahi, A.F.; Khalaf, O.I.; Alotaibi, Y.; Natarajan, R.; Mahadev, N.; Ramesh, T. Modified Self-Adaptive Bayesian Algorithm for Smart Heart Disease Prediction in IoT System. Sustainability 2022, 14, 14208.

DOI: 10.48175/IJARSCT-24974