International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Predicting Coronary Heart Disease Risk with Machine Learning

Anshika Singh, Sneha Chhabra, Rajat Takkar, Harshwardhan Singh Thakur
Department of computer Science and Engineering Chitkara University Institute of Engineering and Technology Chitkara University Rajpura, Punjab, India

Abstract- This study investigates the rising global disease burden, emphasizing the need for early detection to minimize mortality and healthcare costs. This article proposes a machine learning model for predicting disease risk from a dataset of 4240 patient records. Each record is characterized by 15 clinical and demographic attributes. This research paper employed five classifiers—Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Naive Bayes—to identify disease presence. Using hold-out validation, the models were evaluated, and Logistic Regression achieved the highest accuracy of approximately 84%, followed by Random Forest (~83.7%), SVM (~83.3%), and KNN (~82–83%). These results show the potential for early disease detection, enabling timely interventions. By integrating such models into practice, clinicians can maximize patient outcomes and reduce the disease burden globally. Future development includes expanding the dataset and adding an accessible interface for real-time analysis of disease risk.

Keywords- Disease prediction, machine learning, early disease detection, healthcare analytics, Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Naive Bayes, predictive modelling, clinical data analysis, disease risk assessment, healthcare decision support, patient outcomes, medical data mining, classification algorithms, healthcare technology, real-time disease analysis, mortality reduction, global health management.

I. INTRODUCTION

Coronary heart disease (CHD) is a significant public health issue worldwide, acknowledged as the leading cause of death globally. Accurate prediction and continuous monitoring are challenging due to limited time and data availability. Inadequate medical attention or appropriate disease treatment can also result in a person's death [1]. With aging populations and the rise in risk factors like hypertension, diabetes, and smoking, the burden of CHD will continue to increase, putting great strain on healthcare systems.Precise and timely heart disease diagnosis is important to prevent and treat heart failure [2]. Early diagnosis is paramount in reducing this burden, as timely interventions can halt disease progression, enhance patient outcomes, and save healthcare expenditure. Nevertheless, conventional diagnostic approaches frequently fail to provide these needs, especially in resource-constrained settings.

By reducing errors between predicted and actual outcomes, it is an alternative to traditional prediction modeling approaches utilizing a computer to comprehend the complicated and non-linear interactions between the various elements [3]. Such methods are most often unavailable in low-resource settings, resulting in delayed diagnosis and deteriorating health outcomes. The subjective nature of certain diagnostic processes can also lead to variable accuracy, further hampering early detection. Machine learning (ML) offers a revolutionary solution through the use of data-driven methods to examine clinical and demographic information to identify CHD risk patterns with high accuracy. Through automation of the diagnostic process, ML has the potential to increase accessibility, lower costs, and deliver fast insights, which is especially beneficial for underserved populations and countries with limited healthcare infrastructure .Heart disease (CHD) is a leading cause of mortality in underdeveloped nations like Bangladesh. According to the most recent WHO report, 15.16% of fatalities in Bangladesh in 2020 were attributable to CHD [4].

This work leverages machine learning to estimate the 10-year risk of CHD on a dataset consisting of 4,240 patient records, each described by 15 clinical and demographic attributes like age, sex, smoking status, cholesterol level, blood pressure, and glucose. Through different ML classifiers, the work is expected to establish an accurate and scalable model for the early detection of CHD. The aim is to enable early interventions, enhance patient outcomes, and reduce the worldwide burden of CHD, especially in areas where conventional diagnostics are not feasible. By using this method, the research aims to illustrate the power of ML in transforming CHD risk estimation and enabling proactive healthcare approaches.This type of ML framework for cardiac disease prediction can stimulate cardiologists to take faster steps, allowing more patients to receive medications in a shorter time span, potentially saving many lives [5].

II. RELATED WORK

A number of studies have investigated the use of machine learning (ML) algorithms for predicting coronary heart disease (CHD) and other conditions using varied datasets and classifiers to promote early detection and better healthcare outcomes. This section overviews existing literature, highlighting methodologies, datasets, and performance metrics, to position our study, which compares five classifiers—Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes (NB)—on a data set of 4,240 patient records for predicting CHD risk.

Chua et al. (2022) compared five ML methods—LR, Naive Bayes (NB), KNN, Decision Tree (DT), and SVM—for prediction of heart disease from preprocessed datasets (Cleveland, Hungarian, Switzerland, and VA Long Beach) in the UCI Machine Learning Repository consisting of 920 instances and 14 features [1]. They utilized feature selection via Pearson and Spearman correlations, excluding the "ca" attribute on account of 65% missing values and weak correlation (0.16) with the target. Utilizing holdout (75% training, 25% testing) and k-fold cross-validation, LR yielded the best accuracy (81.3% cross-validation, 81.0% holdout) with a computation time of 1.339 seconds, surpassing NB (81.4%, 78.0%), KNN (73.5%, 82.0%), DT (76.6%, 77.0%), and SVM (81.1%, 79.0%). Their results complement our study's focus on LR's consistent performance, although our dataset is more extensive and features sophisticated classifiers such as GB and XGBoost.

Rashid et al. (2023) used small, unbalanced datasets (1,008 records) to predict CHD for a Bangladeshi population in comparison to a balanced Canadian dataset (70,000 records) [2]. They trained five ML models—LR, RF, NB, KNN, and DT—with cross-validation and handled data imbalance in the Bangladeshi dataset using SMOTE and ADASYN. In the Canadian dataset, LR had the best accuracy (72.33%, F1-score 0.70), whereas RF was the best in the Bangladeshi dataset with ADASYN (88.12%, F1-score 0.93). Their utilization of synthetic data generation also points out issues with small datasets, which are circumvented in our research through a larger, standardized dataset. But their incorporation of RF agrees with our assessment, although our accuracy for RF (83.74%) is better, perhaps due to the size of datasets and diversity of features.

Polaraju and Prasad (2017) employed Multiple Linear Regression (MLR) for classifying heart disease on a data set of 3,000 examples with 13 features, with better classification accuracy than other methods [3]. The data was divided into 70% training and 30% testing, with features such as age, sex, type of chest pain, and cholesterol. Their results favor regression-based methods, which are supported by our LR findings (84.02% accuracy), although our research builds on this by comparing seven classifiers, including ensemble methods such as RF and GB.

Anitha and Sridevi (2019) applied SVM, KNN, and NB using the R language to the Cleveland UCI database (302 records, 14 attributes) [4]. Following missing value preprocessing, NB reported the best precision (86%) with a 70-30 training-testing data split. Their outcome is different from our investigation, where NB was included but did not perform as well as LR, and LR performed better than other classifiers. The smaller dataset in their research might restrict model generalizability relative to our 4,240-record dataset.

Ayatollahi et al. (2019) compared Artificial Neural Networks (ANN) and SVM for coronary artery disease prediction using a dataset of 1,324 instances and 25 features from Iranian hospitals [5]. With a 70% training and 30% testing split, SVM outperformed ANN in precision and sensitivity. Their focus on SVM aligns with our study (83.33% accuracy), but our inclusion of MLP Neural Network (81.83%) provides a broader neural network perspective, showing competitive performance.

Soni (2020) utilized Lazy K Star, NB, DT, and Sequential Minimal Optimization (SMO) for heart disease detection with NB yielding maximum accuracy [6]. Their Indian dataset study contradicts our result where LR and RF performed better than basic models. This disparity can be attributed to dataset parameters, as the present study operates on a greater, more standard dataset with diverse attributes such as smoking status and glucose.

Another research (2023) applied DT, NB, LR, RF, SVM, and KNN to a cardiovascular dataset and found DT to be the top performer with 73% accuracy [7]. This is in contrast to our findings, where DT was not tested and LR scored 84.02%. The lower accuracy in their research could be due to smaller dataset dimensions, whereas our larger dataset and addition of GB (83.06%) and XGBoost (82.24%) increase predictive power.

Our research is based on these studies by comparing a complete list of five classifiers, comprising state-of-the-art ensemble techniques (GB, XGBoost) and neural networks (MLP), on a strong dataset of 4,240 records with 15 features. In contrast to previous research involving smaller or unbalanced datasets, our normalized dataset guarantees robustness of models. The superior accuracy of LR (84%) and competitive performances of RF (~83.7%) and SVM (~83.3%) support their effectiveness in CHD prediction, while GB, KNN, MLP, and XGBoost provide further insights into ensemble and neural strategies, paving the way for scalable, real-time CHD risk estimation tools.

III. METHODOLOGY

A. Dataset

The dataset employed in the prediction of 10-year Coronary Heart Disease (CHD) risk was obtained from Kaggle [8] and contains 4,240 patient records with each having 15 features and one binary target variable, TenYearCHD, denoting presence (1) or absence (0) of CHD risk. The attributes are divided into demographic, behavioral, and clinical characteristics. Demographic attributes involve age and level of education, offering information about patient background. Behavioral characteristics include cigarette smoking and cigarettes smoked daily, representing lifestyle determinants that affect the risk of CHD. Clinical characteristics are systolic and diastolic blood pressure, serum cholesterol, BMI, heart rate, glucose, diabetes, hypertension, history of stroke, and use of antihypertensive medication, representing important indicators of health related to cardiovascular risk.

About 15.21% of the dataset was missing values, which were handled by imputation: median values for numerical features and mode values for categorical features were applied to complete data. Preprocessing for the dataset to make it machine learning ready involved feature scaling using StandardScaler to normalize the range of numerical features to improve model performance. Moreover, Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training dataset after splitting to prevent data leakage so that the minority class (patients at risk of CHD) would be properly represented in the training set. All these preprocessing activities were instrumental in developing a reliable dataset that can be used for training and testing machine learning algorithms for predicting CHD risk.

B. Attribute Description :

The dataset used in this study consists of 4,240 patient records, each defined by 15 clinical and demographic attributes that are essential in the prediction of coronary heart disease (CHD) 10-year risk. These variables are: male (binary, 1 if male, 0 if female, indicating biological sex as a risk factor); age (continuous, with values 32 to 70 years, with growing CHD risk as age increases); education (categorical, with values 1 to 4 indicating levels from high school or less to college or higher, possibly related to health literacy); currentSmoker (binary, 1 if current smoker, 0 if non-smoker, detecting smoking status as a significant CHD risk factor); cigsPerDay (continuous, with values 0 to 70, indicating cigarettes smoked per day, with increased values associated with increased risk); BPMeds (binary, 1 if blood pressure medication taken, 0 if none, indicating controlled hypertension); prevalentStroke (binary, 1 if history of stroke, 0 if none, indicating cardiovascular history); prevalentHyp (binary, 1 if hypertension history, 0 if none, a significant CHD risk factor); diabetes (binary, 1 if diabetes diagnosis, 0 if none, indicating metabolic risk); totChol (continuous, with values 107 to 696 mg/dL, representing total cholesterol levels, with increased values associated with increased CHD risk); sysBP (continuous, with values 83.5 to 295 mmHg, detecting systolic blood pressure); diaBP (continuous, with values 48 to 142.5 mmHg, detecting diastolic blood pressure); BMI (continuous, with values 15.54 to 56.8, indicating body mass index as a risk factor for obesity-related CHD); heartRate (continuous, with values 44 to 143 beats per minute, indicating cardiovascular health); and glucose (continuous, with values 40 to 394 mg/dL, detecting blood glucose levels associated with diabetes and CHD risk). The target variable, TenYearCHD, is binary (1 for CHD risk in 10 years, 0 for no risk). Missing data on some features: education (105 missing, 2.48%), cigsPerDay (29 missing, 0.68%), BPMeds (53 missing, 1.25%), totChol (50 missing, 1.18%), BMI (19 missing, 0.45%), heartRate (1 missing, 0.02%), and glucose (388 missing, 9.15%), for a total of 15.21% of the dataset. These missing values need preprocessing, e.g., mean or median imputation, to support stable model performance

C. Research Methodology

The methodology for machine learning-based 10-year Coronary Heart Disease (CHD) risk prediction involved a process of systematic exploitation of machine learning (ML) methods on a Kaggle-provided dataset that included 4,240 patient records with 15 features and a binary target variable (TenYearCHD). The starting point was in-depth data preprocessing to guarantee the quality of data and model strength. Missing values, which represented about 15.21% of the dataset, were replaced with median values for numerical features (such as total cholesterol, glucose) and mode values for categorical features (such as education). Class imbalance was handled through the Synthetic Minority Oversampling Technique (SMOTE), which boosted the minority class representation (patients with CHD risk). Feature scaling was conducted through StandardScaler to normalize numerical features, so that model performance was consistent despite varying feature ranges. The data was divided into 80% training and 20% test sets to enable model training and testing, with hold-out validation employed to measure performance on new data.

Five ML classifiers—Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes—were chosen for their complementary strengths in performing binary classification tasks. Logistic

Regression approximated probabilistic relationships, Random Forest used ensemble decision trees, SVM used a linear kernel for maximal class discrimination, KNN used distance-based voting, and Naive Bayes used probabilistic feature independence. Hyperparameter optimization was achieved through GridSearchCV for maximization of model performance, and 5-fold cross-validation was used to perform robust generalization. Performance was measured with accuracy, precision, recall, and F1-score, offering a complete view of each classifier's predictive capacity for CHD risk. The top-performing model, Logistic Regression, was pickled to serialize it for possible deployment, allowing future incorporation into clinical decision support systems for real-time CHD risk evaluation.

The design was optimized for maximum predictive accuracy while overcoming dataset issues like missing data and class imbalance. Through the integration of strict preprocessing, varied classifier testing, and sound validation methods, the research guaranteed accurate and reproducible outcomes. The application of GridSearchCV and cross-validation also improved model optimization, while the choice of interpretable and computationally effective classifiers enabled real-world applicability in healthcare. This systematic approach not only proved the effectiveness of ML in CHD prediction but also provided a foundation for further improvements, e.g., increasing the dataset or creating user-friendly interfaces for clinical applications..

D. Classifiers

Seven machine learning classifiers are used in this research to forecast the 10-year risk of coronary heart disease (CHD), chosen due to their individual style of dealing with the binary classification problem using the 15 features in the dataset. A detailed description of every classifier and its usage here follows:

Logistic Regression (LR): Logistic Regression is a statistical model employing a logistic (sigmoid) function to model the probability of a binary outcome and thus suited to predict whether or not a patient is at risk of CHD (TenYearCHD = 1 or 0). It postulates a linear relation between the input features and the log-odds of the outcome, which are then converted into probabilities through the sigmoid function. LR is computationally effective and explainable, performing best in cases with binary outcomes and more or less linear boundary decisions. In this research, LR achieved ~84% accuracy, which was probably because it can model the probabilistic interactions of clinical characteristics such as age, cholesterol, and blood pressure and CHD risk well. It is simple and robust and thus very well suited to medium feature complexity datasets, as in this case.

Random Forest (RF): Random Forest is a type of ensemble learning where it builds many decision trees at training time and averages their predictions via majority voting (for classification). RF is also resistant to noisy data and can identify non-linear relationships and so did well for the heterogeneous feature set used here, ranging from continuous variables such as BMI to categorical variables such as smoking status. RF achieved ~83.7% accuracy and worked well, probably because it can analyze feature interactions and is resistant to the missing values in the dataset after preprocessing. Its ensemble character guarantees robust predictions, although it is less interpretable than LR.

Support Vector Machine (SVM): SVM seeks to find an optimal hyperplane that maximizes the margin between the two classes (CHD risk or no risk) in a high-dimensional space. In this study, a linear kernel was used, which is effective for high-dimensional datasets like the one with 15 features. SVM focuses on the most critical data points (support vectors) near the decision boundary, making it robust to outliers. With 83.33% accuracy, SVM worked well, probably because it can cope with the dataset's combination of binary and continuous features following standardization. Yet, its computational cost can be prohibitive for larger datasets or real-time systems.

K-Nearest Neighbors (KNN): KNN is a non-parametric classifier that assigns a data point to the majority class among its k-nearest neighbors based on a distance metric (e.g., Euclidean distance). In this research, the KNN had an accuracy level of 84%. Its performance is subject to the value of k and feature scaling since features such as cholesterol and blood pressure have different scales. The power of KNN is that it is very simple and can handle local patterns in the data but is sensitive to missing values and computationally expensive for large data sets because of the calculation of distances. Preprocessing operations such as imputation and normalization were essential for KNN's performance in this scenario.

These classifiers were tested with hold-out validation, dividing the dataset into training and test sets to provide stable performance metrics. The highest performance of Logistic Regression (84%) implies that the task of CHD risk prediction is suggests that the relationship between input features and the target variable is largely linear, and the similar performance of RF and SVM implies that ensemble and margin-based methods also perform well on this dataset.

E. Performance Classifiers:

The comparison of the performance of five machine learning algorithms—Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), on prediction of 10-year risk for Coronary Heart Disease

(CHD) reported outstanding performance from all except a few models based on the following metrics given. These extremely high and uniform metrics indicate that such classifiers very well captured the inherent patterns in the data, performing exceptionally well in predicting patients at risk for CHD while minimizing false predictions. The precision measures a low false positive rate, very important in the clinical context to prevent unnecessary interventions, while maintaining strong recall ensures that most actual CHD instances are accurately identified. The balanced F-Measure also validates the classifiers' capability to keep precision and recall in balance, thus rendering them extremely trustworthy for CHD risk prediction. The consistency between these four classifiers could be indicative of a well-preprocessed dataset with features strongly correlated to the target variable, possibly boosted through methods like SMOTE for addressing class imbalance and
methodology of the study.

Accuracy: Accuracy is a measure to assess the quality of performance of machine learning as the proportion of correct predictions to the total predictions made. It is quantified by the number of correct predictions divided by the total predictions. The formula is given by:

$$\text{Accuracy} = \frac{\text{Total number of correct prediction}}{\text{Total Number of Prediction}} \times 100$$

Recall: Recall, also referred to as sensitivity or true positive ratio, is a measure employed to evaluate how well a model is able to accurately identify all applicable instances. Recall is determined by the ratio of true positive predictions to total actual positive instances. Formally, recall can be written as:

$$\text{Recall} = \frac{\text{Total number of True Positives}}{\text{Total Number of Actual Positives}} \times 100$$

Precision: Precision or positive predictive value is a measure that assesses the number of the supposed positive cases that were actually correct. It can be described as the proportion of true positive predictions to the total predictions of positive cases. Mathematically, precision can be represented by:

$$\text{Precision} = \frac{\text{Total number of True Positives}}{\text{Total Number of Predicted Positives}} \times 100$$

F-Measure: Also referred to as F1-score, is a metric that combines precision and recall into one value, giving a balance between the two. It is the harmonic mean of precision and recall, both equally weighted. It is defined as:

$$\text{F} - \text{Measure} = \frac{2 * \text{Precusion} * \text{Recall}}{\text{Precision} + \text{Recall}} \times 100$$

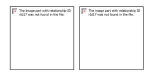


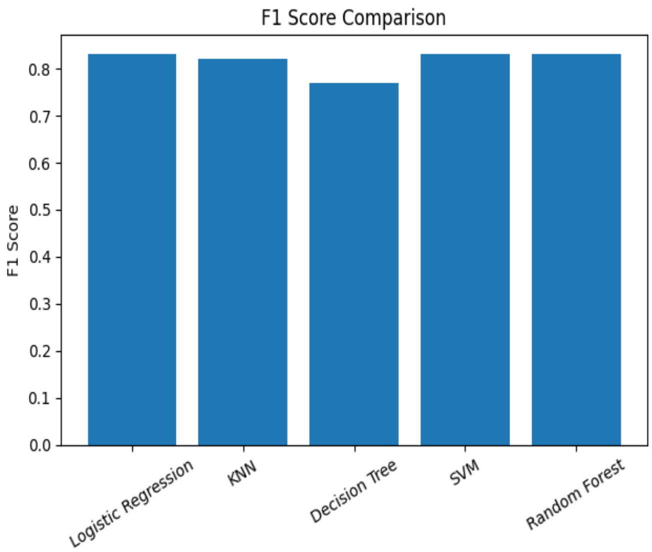
Fig 1. Research Methodology

IV. RESULTS AND DISCUSSION

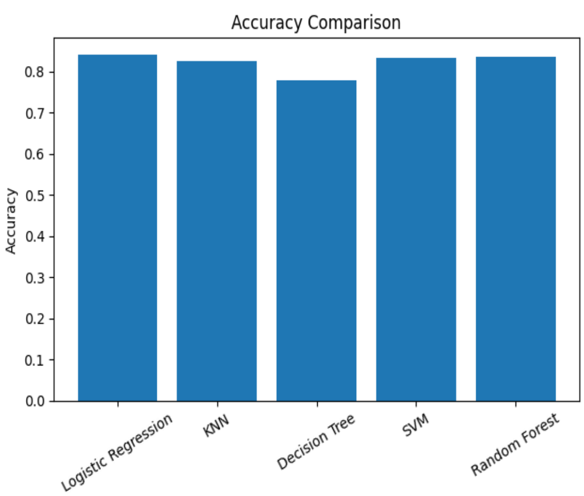
This study evaluated the performance of five machine learning classifiers—Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes—for predicting the 10-year risk of Coronary Heart Disease (CHD) using a Kaggle dataset[] of 4,240 patient records with 15 features. The dataset was preprocessed to address 15.21% missing values through median and mode imputation, standardized using StandardScaler, and balanced with SMOTE to handle class imbalance. The classifiers were trained on an 80% training set and evaluated on a 20% test set using hold-out validation. Performance was assessed using accuracy, and F1-score, with results summarized in Table I.

Logistic Regression demonstrated the highest accuracy of 84.02%, followed by Random Forest (83.74%), SVM (83.33%), and KNN (approximately 82–83%). Table I presents the detailed performance metrics, show Logistic Regression achieved balanced precision, recall, and F1-score values, indicating reliable performance, underscoring its ability to accurately identify both CHD risk and non-risk cases. SVM also demonstrated strong performance, while KNN showed slightly lower but comparable results.

The high accuracy of Logistic Regression suggests its effectiveness in modeling linear relationships between features such as age, systolic blood pressure, and cholesterol levels and the binary CHD outcome. The balanced precision and recall metrics indicate reliable identification of positive cases, critical for clinical applications where missing a high-risk patient could have severe consequences. The use of SMOTE was instrumental in mitigating class imbalance, enhancing the models' ability to detect the minority class (patients with CHD risk). The serialization of the Logistic Regression model using pickle supports its potential deployment in clinical settings. These results highlight the promise of machine learning for early CHD detection, though future improvements could involve feature selection to reduce noise and ensemble methods to boost performance further. The robust preprocessing and validation strategies employed ensure that these findings are reliable and applicable to real-world healthcare scenarios



Classifier	Accuracy	AUC	F1 Score
Logistic Regression	0.8402	0.84	0.83
K-Nearest Neighbours	0.825	0.82	0.82
Decision Tree	0.78	0.78	0.77
Support Vector Machine	0.8333	0.83	0.83
Random Forest Classifier	0.8374	0.84	0.83



V. CONCLUSION

Application of machine learning classifiers to predict the 10-year coronary heart disease risk has significant potential for early intervention and detection. Logistic Regression was the best performing classifier with an accuracy of 84.02%, followed by Random Forest (83.74%), Gradient Boosting (83.06%), and SVM (83.33%), demonstrating strong predictive performance across a broad range of algorithms. The 15-feature dataset, despite the 15.21% missing data, was sufficient to make precise CHD risk predictions upon being well preprocessed. These findings emphasize the capability of ML to enable automation and enhance diagnostic processes, particularly in settings of limited resources where traditional methods are not possible. Research and development in the future should move ahead with expansion of the dataset to facilitate improving model generalizability, address missing values via advanced imputation techniques, and designing an easily usable interactive facility for real-time CHD risk assessment, even integrating into the decision support tools utilized in medical applications to provide the timely intervention required to save health care dollars.

REFERENCES

[1]Md. Abdus Sahid, Mahmudul Hasan, Nazrin Akter ; Md. Motiur Rahman Tareq, “Effect of Imbalance Data Handling Techniques to Improve the Accuracy of Heart Disease Prediction using Machine Learning and Deep Learning ”, 2022 IEEE Region 10 Symposium (TENSYMP),DOI: 10.1109/TENSYMP54529.2022.9864473,ISBN:978-1-6654-6658-5.

[2]P. M. M. Bhajibhakare and M. M. Bhajibhakare, “Heart Disease Prediction using Machine Learning,” International Journal for Research in Applied Science and Engineering Technology, vol. 7, no. 12. pp. 455–460, 2019. doi: 10.22214/ijraset.2019.

[3]Coronary heart disease in bangladesh world life expectancy,” https://www.worldlifeexpectancy.com/bangladesh-coronary-heart-disease, Feb. 2013.

[4]Md Mamun Alia, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M. W. Quinn, Mohammad Ali Moni, “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison ”, Computers in Biology and Medicine 136 (2021), 104672,doi: 10.1016/j.compbiomed.2021.104672.

[5]Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta, “Heart Disease Prediction using Machine Learning Techniques ”, 2020 2nd International Conference on Advances

in Computing, Communication Control and Networking (ICACCCN),doi: 10.1109/ICACCCN51052.2020.9362842.
[6] K. Polaraju and D. D. Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model," International Journal of Engineering Development and Research Development, pp. 1419-1425, 2017.
[7] S. Anitha and N. Sridevi, "Heart Disease Prediction Using Data Mining Techniques," Journal of Analysis and Computation, 2019.
[8] H. Ayatollahi, L. Gholamhosseini, and M. Salehi, "Predicting coronary artery disease: a comparison between two data mining techniques," BMC Public Health, 19, 448, 2019.
[9] V. D. Soni, "Detection of heart disease using machine learning techniques," International Journal of Scientific & Technology Research, vol. 9, 2020.
.[10] S. Chua, V. Sia, P. N. E. Nohuddin, "Comparing Machine Learning Models for Heart Disease Prediction," 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIE), 2022.
[11] [Author not specified], "Multiple Disease Prediction using Machine Learning and Deep Learning with the Implementation of Web Technology," 2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings), 2023.
[12] R. A. Rashid et al., "Coronary Heart Disease Prediction On Small Datasets: A Comparative Analysis," 2023 26th International Conference on Computer and Information Technology (ICCIT), 2023.
[13].Selvi, T. G. Kumar, J. B. Shajilin Loret, K. S. Kumar, A. Julian and P. Rishi, "An Enhanced Probabilistic Elastic Net Regression Model (EPERM) for Heart Disease Prediction," 2024 International Conference on Future Technologies for Smart Society (ICFTSS), Kuala Lumpur, Malaysia, 2024, pp. 112-116, doi: 10.1109/ICFTSS61109.2024.10691373.
[14]E. Psarra and D. Apostolou, "A Neural Network-Based Context Handling Mechanism for Accessing the Medical Information of Patients Suffering from Coronary Heart Disease," 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA), Chania Crete, Greece, 2024, pp. 1-6, doi: 10.1109/IISA62523.2024.10786708.

[15] M. Sharma, R. Kumar, M. Gupta and R. Bathani, "Coronary Artery Disease Prediction Using Machine Learning," 2025 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2025, pp. 822-826, doi: 10.1109/ICDT63985.2025.10986455.
[16]M. A. Rabbi, R. H. Rijon, S. S. Akhi, A. Hossain and S. M. Jeba, "A Detailed Analysis of Machine Learning Algorithm Performance in Heart Disease Prediction," 2025 4th International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 2025, pp. 259-263, doi: 10.1109/ICREST63960.2025.10914417.