iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: III Month of publication: March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58797

Revolutionizing Cardiovascular Health: A Machine Learning Approach for Predictive Analysis and Personalized Intervention in Heart Disease

Shilpy Agrawal

Abstract: Cardiovascular Diseases (CVDs) continue to be a leading cause of global morbidity and mortality, necessitating innovative approaches for early detection and personalized interventions. This research explores the transformative potential of machine learning (ML) in revolutionizing cardiovascular health. Leveraging advanced predictive analytics, our study employs a comprehensive dataset of cardiovascular parameters, incorporating clinical, genetic, and lifestyle factors. The machine learning model developed demonstrates remarkable accuracy in predicting the risk of heart disease, enabling early identification of individuals susceptible to CVD. This predictive analysis empowers healthcare professionals with a powerful tool for pre-emptive intervention and tailored treatment strategies. By considering individual variations in genetic predispositions, lifestyle choices, and clinical data, our approach moves beyond traditional risk assessment models, paving the way for a more personalized and effective healthcare paradigm. Furthermore, the integration of real-time monitoring devices and continuous data streams enhances the adaptability and responsiveness of our model. This allows for dynamic adjustments in treatment plans, ensuring ongoing optimization based on the evolving health status of each patient. The synergy between machine learning and cardiovascular health not only augments diagnostic precision but also facilitates a proactive healthcare ecosystem that prioritizes preventive measures.
Keywords: Machine Learning, Predictive Analysis, Cardiovascular Disease, Heart Disease, Healthcare advancements

I. INTRODUCTION

Cardiovascular diseases (CVDs) stand as a formidable global health challenge, contributing substantially to morbidity and mortality rates across diverse populations. Despite significant advancements in medical science, the intricate nature of heart-related ailments demands a paradigm shift towards more precise and personalized healthcare strategies. In this pursuit, machine learning (ML) emerges as a transformative tool, offering unprecedented potential for predictive analysis and tailored interventions in the realm of cardiovascular health. As we delve into the 21st century, the integration of advanced technologies becomes imperative in the quest for more effective healthcare solutions. Traditional risk assessment models, while valuable, often lack the granularity necessary to address the intricate interplay of genetic predispositions, lifestyle choices, and clinical indicators that contribute to an individual's cardiovascular health profile[1]. The limitations of conventional approaches underscore the need for innovative methodologies capable of capturing the multifaceted nature of heart disease.

This research Endeavor to explore and elucidate the synergies between machine learning and cardiovascular health, with a focus on predictive analysis and personalized interventions. By harnessing the power of comprehensive datasets encompassing diverse parameters, ranging from genetic markers to lifestyle habits, we aim to develop a robust ML model capable of predicting the risk of heart disease with unprecedented accuracy[2]. Such predictive analytics lay the foundation for the early identification of individuals susceptible to CVD, paving the way for timely and targeted interventions. Moreover, the advent of real-time monitoring devices and the integration of continuous data streams offer a dynamic dimension to cardiovascular healthcare. This real-time feedback loop enhances the adaptability of our ML model, allowing for continuous refinement of risk predictions and intervention strategies based on the evolving health status of each individual.

In essence, this research sets out to redefine the landscape of cardiovascular health by harnessing the capabilities of machine learning. Through the lens of predictive analysis and personalized interventions, we aim to revolutionize the approach to heart disease, ushering in an era where healthcare is not merely reactive but proactive, where interventions are not generic but tailored to the unique needs of each patient. The convergence of machine learning and cardiovascular health holds the promise of a future where precision and personalization guide Endeavor to mitigate the impact of cardiovascular diseases on a global scale.

II. WHAT IS CARDIOVASCULAR DISEASE?

Cardiovascular disease (CVD) refers to a class of disorders affecting the heart and blood vessels, compromising their structure and function. It encompasses a range of conditions, including coronary artery disease, heart failure, valvular diseases, and vascular diseases, all of which contribute significantly to global morbidity and mortality[3].

A. Prevalence and Impact

CVD is a pervasive health concern worldwide, responsible for a substantial portion of premature deaths. Factors such as aging populations, sedentary lifestyles, poor dietary habits, and increasing rates of diabetes contribute to the escalating prevalence of cardiovascular diseases. The impact extends beyond mortality, affecting the quality of life and placing a significant economic burden on healthcare systems.

B. Risk Factors

Numerous risk factors contribute to the development of cardiovascular diseases. These include modifiable factors such as smoking, high blood pressure, high cholesterol levels, diabetes, and obesity, as well as non-modifiable factors like age, gender, and genetics[1]. The interplay of these factors increases the likelihood of developing CVD.

C. Pathophysiology

The pathophysiology of cardiovascular disease varies depending on the specific condition. Coronary artery disease, for example, involves the gradual buildup of plaques in the coronary arteries, leading to reduced blood flow to the heart muscle. Heart failure results from the heart's inability to pump blood effectively. Valvular diseases affect the heart valves, compromising their ability to regulate blood flow. Understanding the underlying mechanisms is crucial for diagnosis and treatment[13].

D. Clinical Presentation

Cardiovascular diseases manifest with a diverse range of symptoms. Chest pain, shortness of breath, fatigue, palpitations, and swelling are common indicators. However, the presentation can be subtle or asymptomatic, emphasizing the importance of regular health check-ups and screenings for early detection.

E. Diagnostic Methods

Diagnostic approaches for cardiovascular diseases include non-invasive methods such as electrocardiography (ECG or EKG), echocardiography, and imaging techniques like CT angiography and MRI. Invasive procedures like cardiac catheterization may be employed for a more detailed assessment.

F. Treatment and Management

Management strategies for cardiovascular diseases encompass lifestyle modifications (e.g., diet, exercise, smoking cessation), medications (e.g., anti-hypertensives, statins), and interventional procedures (e.g., angioplasty, bypass surgery). The choice of treatment depends on the specific condition and its severity[14].

G. Preventive Measures

Preventive measures play a crucial role in reducing the burden of cardiovascular diseases. Public health initiatives promoting healthy lifestyles, awareness campaigns, and early screening contribute to prevention. Education on risk factor modification and adherence to prescribed treatments are essential components of cardiovascular disease prevention[15].

III.METHODOLOGY

A. Data Collection

The CVD data set used for developing the detection models was taken from the Kaggle and has been converted into a.csv comma-separated file. It contains 3390 samples and 17 attributes.

Only 10 important test attributes (age, sex, BP, BMI, cholesterol, smoking, and one target output (1 = patient having CVD, 0 = patient not having CVD) have been considered out of the 17 attributes to train and test the model. Our target value taken is whether a person has CVD (near 1) or does not have CVD (close to 0). The data set was imbalanced as 2879 patients had no CVD and 511 patients had CVD.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

B. Data Preprocessing

Data preprocessing is a crucial step in machine learning (ML) that involves cleaning and transforming raw data into a format suitable for training models. This process is essential to ensure the quality and reliability of the data used to train ML algorithms[4]. The goal of data preprocessing is to handle missing values, remove outliers, and standardize or normalize features to make them comparable. Cleaning tasks may include dealing with duplicate records, correcting errors, and converting categorical variables into numerical formats through techniques like one-hot encoding. Additionally, scaling features to a common range helps prevent certain features from dominating the learning process. Data preprocessing aims to enhance the performance and accuracy of machine learning models by addressing issues related to the quality, structure, and format of the input data, ultimately contributing to the effectiveness of the overall learning process[5].

C. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) plays a crucial role in the field of machine learning (ML) by helping practitioners understand and gain insights from their datasets. EDA involves the use of statistical and visual methods to explore data, identify patterns, and detect anomalies before building predictive models. In the context of ML, EDA serves as a preliminary step to preprocess and clean data, ensuring that it is suitable for training and testing models. Through various statistical measures and visualizations, practitioners can uncover relationships between variables, detect outliers, and make informed decisions about feature engineering and selection[5]. EDA also aids in identifying potential biases and understanding the distribution of target variables, guiding the selection of appropriate ML algorithms. Overall, the integration of EDA in the ML workflow enhances the robustness and effectiveness of models by providing a comprehensive understanding of the underlying data dynamics.

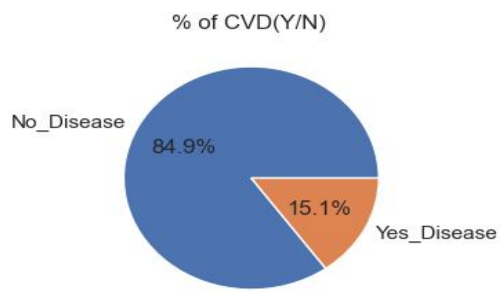

Figure 1. Target Variable Patient is having a disease or not

Figure 1, Our Target Variable is Ten Year Cardiovascular Disease. According to the dataset, 84.9% of people do not have a disease and 15.1% of people have the disease.

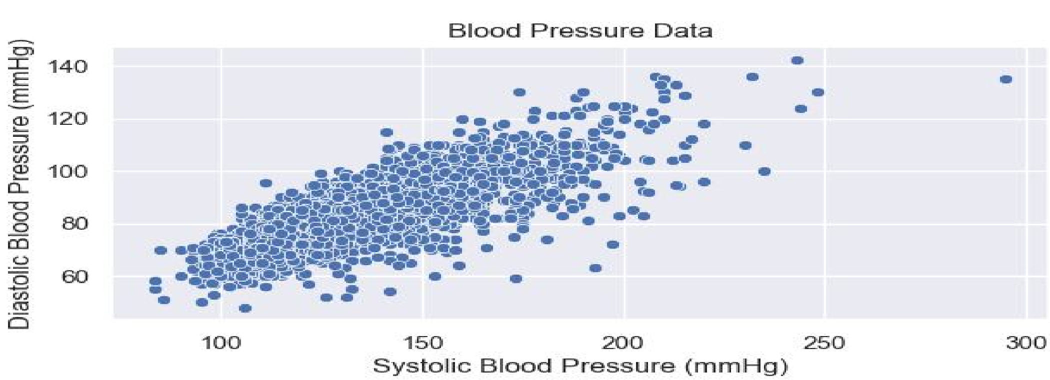

Figure 2. Scatter Plot of Blood Pressure Data

Figure 2, The range of blood pressure mentioned in the dataset is exceedingly out of 250 mmHg. So, the values of systolic pressure (sysBP) above 40 mmHg and below 300 mmHg, whereas diastolic pressure (disBP) is above 20 mmHg and below 300 mmHg.

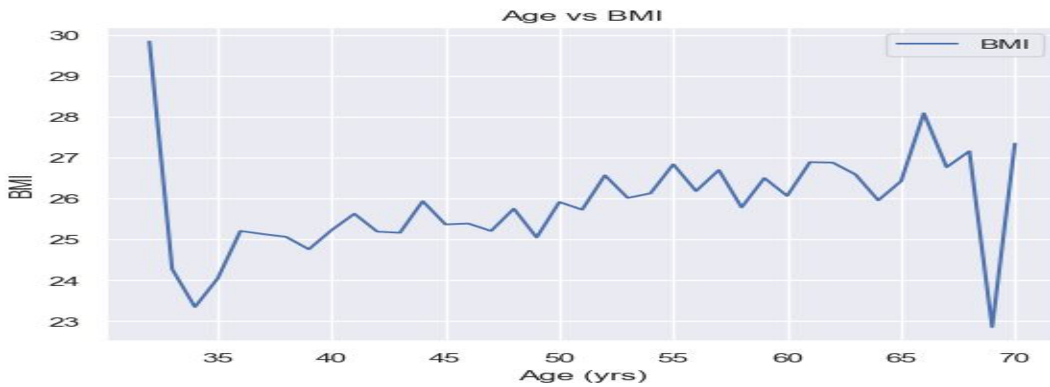

Figure 3. Line Graph of BMI(Body Mass Index) vs Age

Figure 3, The normal range of BMI comes under 18.5 to 25 for a healthy human. The BMI is increasing with age.

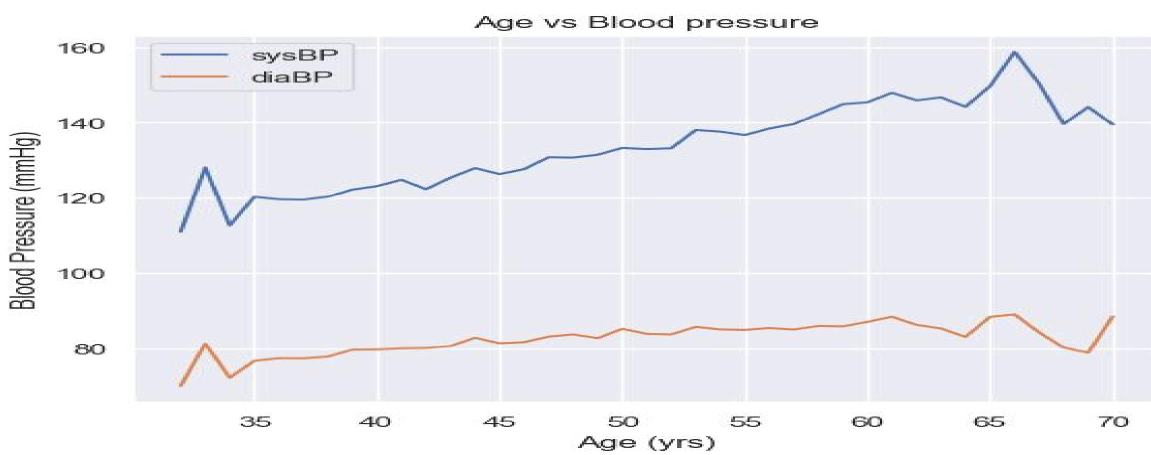

Figure 4. Line Graph of Blood Pressure vs Age

Figure 4, This analysis of age for blood pressure shows that the blood pressure of the patients increases with age. Patients between the ages of 30 to 70 years have an increasing diastolic pressure ranging between 70 to 90 mmHg and increasing systolic pressure ranging between 110 to 160 mmHg.

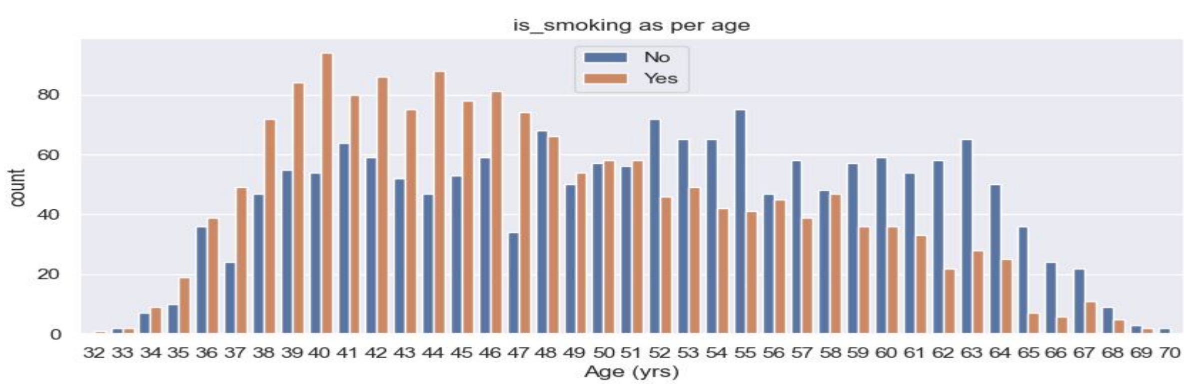

Figure 5. Bar Graph of Smoking as per Age

Figure 5, This analysis shows that most of the patients don't smoke and with an increase in age most of the adults don't prefer to smoke.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

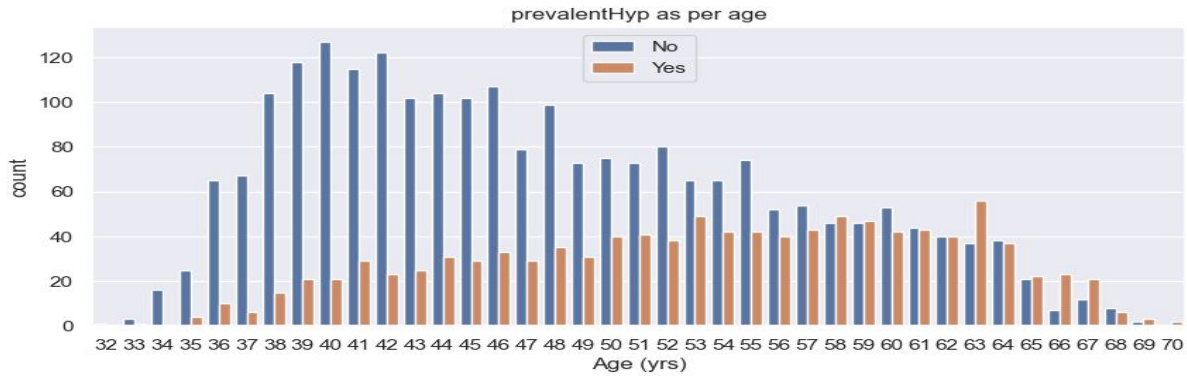

Figure 6. Bar Graph of Prevalent Hypertension as per Age

Figure 6, This analysis shows that hypersensitivity increases with increasing age.

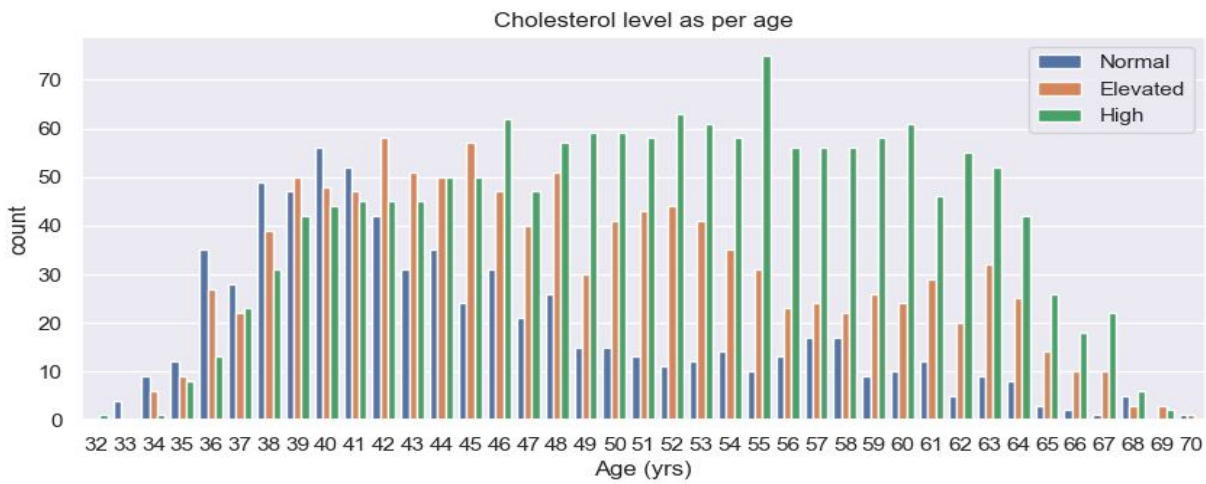

Figure 7. Bar Graph of Cholesterol Level as Per Age

Figure 7, The arteries become narrowed and blood flow to the heart muscle is slowed down or blocked. This can cause chest pain (angina) or a heart attack. That's why high cholesterol levels in adults can cause a serious problem.

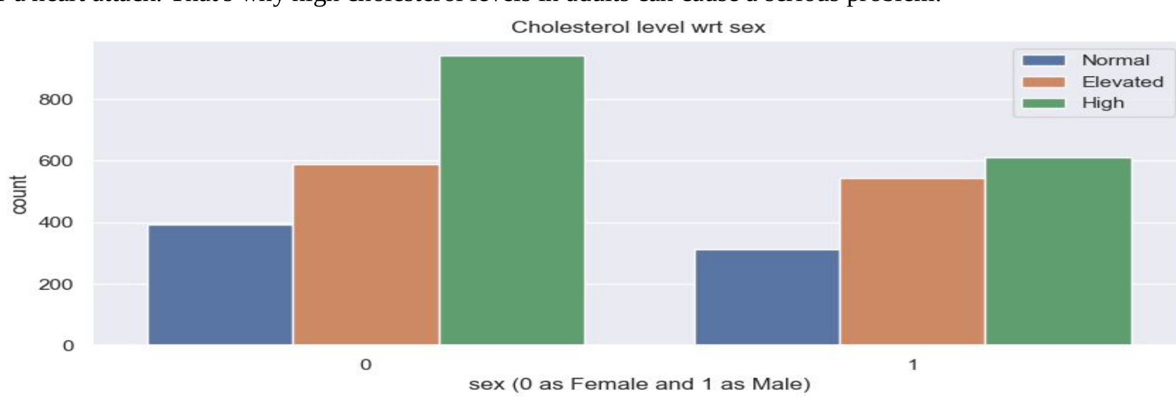

Figure 8. Bar Plot of Cholesterol Level vs Sex

Figure 8, This analysis shows that most of the male and female patients have high cholesterol levels. The data of many female patients present in the dataset is higher than that of male patients, so that's why we cannot compare the cholesterol levels of males and females with each other.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

D. Model Selection
Model selection is a crucial step in machine learning that involves choosing the most appropriate algorithm or model for a given task. It is a critical decision as the performance of the selected model directly impacts the success of the machine learning system. The process typically involves evaluating multiple candidate models based on their ability to accurately generalize patterns from the training data to new, unseen data. Common techniques for model selection include train test split, where the dataset is split into training and testing, and metrics such as accuracy, precision, recall, or F1 score are used to assess performance[6][7]. Additionally, hyperparameter tuning may be performed to optimize the model's configuration for improved performance. The choice of the model should be guided by the specific characteristics of the data, the complexity of the problem, and considerations such as interpretability, computational efficiency, and scalability. Overall, effective model selection is crucial for developing robust and accurate machine-learning solutions.

E. Balance the Imbalance Data
Here data is imbalance data so we need to balance the data to get better accuracy and increase the modal performance using the SMOTE technique. The Synthetic Minority Over-sampling Technique, commonly known as SMOTE, is a valuable method in the field of machine learning, specifically designed to address class imbalance in datasets. Class imbalance occurs when one class significantly outnumbers the other(s), potentially leading machine learning models to Favor the majority class and perform poorly on minority class predictions. SMOTE tackles this issue by oversampling the minority class through the generation of synthetic instances. The SMOTE algorithm works by selecting a minority class instance and creating synthetic examples along the line segments connecting it to its nearest neighbors. By introducing these synthetic instances, SMOTE effectively increases the representation of the minority class, helping the model to learn more robust decision boundaries and preventing it from being biased towards the majority class. One notable advantage of SMOTE is its ability to enhance the predictive performance of classifiers, particularly in scenarios where the minority class is underrepresented. It is widely used in various applications, such as fraud detection, medical diagnosis, and text classification, where imbalanced datasets are common. After balancing the data we had 5758 records and 10 attributes.

F. Training the Model
Split the dataset into training and validation sets for model training and evaluation. Utilize the training set to teach the machine learning model to recognize patterns associated with cardiovascular health.

G. Algorithms
Algorithms in machine learning are systematic sets of rules and mathematical instructions that enable computers to learn patterns and make predictions or decisions without explicit programming. These algorithms serve as the backbone of machine learning models, allowing systems to analyze data, identify trends, and derive insights[12]. These can be categorized into supervised learning, where models are trained on labeled data to make predictions, unsupervised learning, where algorithms uncover patterns in unlabelled data, and reinforcement learning, where agents learn from interacting with an environment to maximize rewards. Common algorithms include decision trees, support vector machines, neural networks, and clustering methods. The selection of a particular algorithm depends on the nature of the task, the characteristics of the data, and the desired outcomes[8], reflecting the diversity and adaptability inherent in machine learning.
Here our data is labelled data so we will use supervised learning algorithms.

1) *Decision Tree Algorithm:* A Decision Tree is a versatile and intuitive machine-learning algorithm used for both classification and regression tasks. It operates by recursively partitioning the dataset into subsets based on the most significant features, creating a tree-like structure where each internal node represents a decision based on a feature, and each leaf node holds the predicted outcome. The tree-building process involves selecting the optimal feature at each step, aiming to maximize information gain or minimize impurity. Decision Trees are transparent and easy to interpret, making them valuable for explaining the reasoning behind predictions. However, they can be prone to overfitting and capturing noise in the data[8]. To address this, techniques such as pruning or ensemble methods like Random Forests are often employed. Despite their simplicity, Decision Trees serve as building blocks for more sophisticated algorithms and find applications in various domains, including finance, healthcare, and natural language processing. For a decision tree, the confusion matrix helps assess the model's accuracy, precision, recall, and F1 score, among other metrics. Precision is the ratio of correctly predicted positive observations to the total predicted positives (TP / (TP + FP)), while recall is the ratio of correctly predicted positive observations to all observations

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

in actual positive class (TP / (TP + FN))[9]. The F1 score is the harmonic mean of precision and recall. Here's how the confusion matrix is related to a decision tree's evaluation:

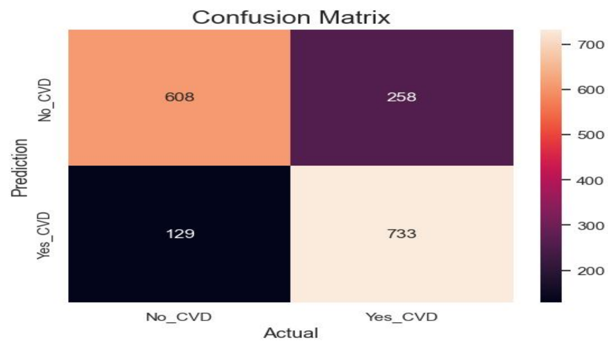

Figure 9. Confusion Matrix for Decision Tree Algorithm

2) *Random Forest Algorithm:* Random Forest is a versatile and powerful machine learning algorithm widely employed for both classification and regression tasks. It belongs to the ensemble learning family, constructing an ensemble of decision trees to enhance predictive accuracy and robustness. The fundamental concept revolves around the aggregation of multiple weak learners, and individual decision trees, in this case, to form a robust and accurate model. During training, each tree is grown on a subset of the training data through a process known as bagging (Bootstrap Aggregating), where random samples with replacement are used to create diverse and independent trees. Additionally, randomness is introduced by selecting a random subset of features at each split in the tree-building process[9]. The final prediction is then determined through a voting mechanism in classification tasks or an averaging process in regression tasks, combining the predictions of all the individual trees. This randomness and aggregation effectively mitigate overfitting and enhance generalization performance, making Random Forest a popular choice in various machine learning applications, especially when dealing with high-dimensional and complex datasets.

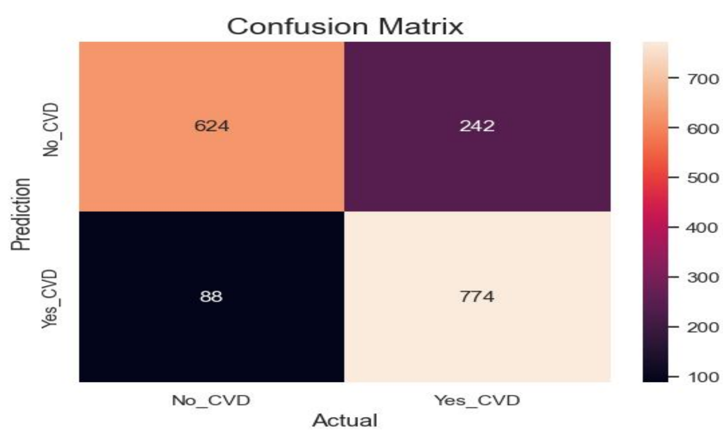

Figure 10. Confusion Matrix for Random Forest Algorithm

3) *Support Vector Machine:* Support Vector Machine (SVM) is a powerful supervised learning algorithm widely used for classification and regression tasks. The fundamental idea behind SVM is to find a hyperplane that best separates different classes in the feature space. This hyperplane is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data points of each class, known as support vectors. SVM is particularly effective in high-dimensional spaces, making it suitable for a variety of applications such as image classification, text categorization, and bioinformatics. One of the strengths of SVM is its ability to handle non-linear decision boundaries through the use of kernel functions, which map the input features into a higher-dimensional space[6]. This flexibility allows SVM to capture complex relationships within the data. SVMs are robust, efficient, and less prone to overfitting, making them a popular choice in machine learning for both linear and non-linear classification problems.

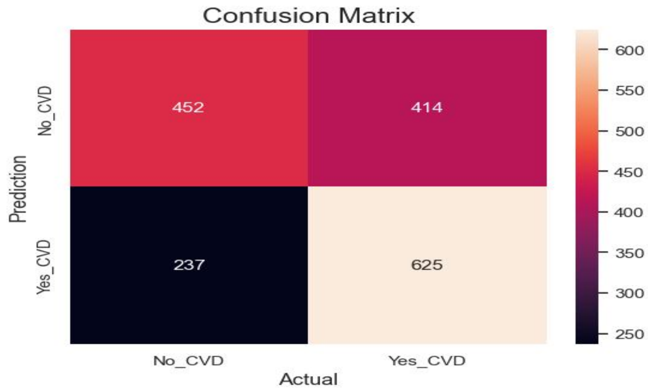

Figure 11. Confusion Matrix for Support Vector Machine Algorithm

4) *K Nearest Neighbour:* K-Nearest Neighbours (KNN) is a simple yet powerful machine learning algorithm employed for both classification and regression tasks. The fundamental idea behind KNN is to make predictions based on the majority class or average output of the k-nearest data points in the feature space. The algorithm operates on the assumption that similar instances in the input space should have similar output values. To determine the proximity of data points, a distance metric, often Euclidean distance, is commonly used. One notable advantage of KNN is its simplicity and ease of implementation. Since the algorithm doesn't make explicit assumptions about the underlying data distribution, it is considered non-parametric. However, the performance of KNN can be sensitive to the choice of the distance metric and the value of k. A smaller value of k can lead to more flexible decision boundaries but may also make the algorithm more susceptible to noise, while a larger k may smooth out local patterns[7]. KNN is particularly suitable for applications where the decision boundary is not easily defined, and the relationships within the data are expected to be complex and local. It finds applications in image recognition, recommendation systems, and cases where the dataset does not have a clear underlying structure. Despite its simplicity, KNN remains a valuable tool in the machine learning toolkit, especially in situations where interpretability and adaptability to local patterns are crucial.

Figure 12. Confusion Matrix for KNN Algorithm

5) *Logistic Regression:* Logistic Regression is a widely used statistical method and classification algorithm in the field of machine learning. Despite its name, logistic regression is employed for binary classification tasks, where the goal is to predict the probability of an instance belonging to one of two classes. This algorithm is particularly well-suited for scenarios where the relationship between the features and the binary outcome can be modeled linearly. In logistic regression, the logistic function (also known as the sigmoid function) is employed to transform a linear combination of input features into a probability score bounded between 0 and 1[2]. This probability score represents the likelihood of an instance belonging to the positive class. The logistic function's S-shaped curve ensures that extreme values are avoided, making it convenient for probabilistic interpretation. Logistic regression offers several advantages, including simplicity, interpretability, and efficiency. It performs well when the relationship between the features and the outcome is approximately linear. Additionally, logistic regression provides probabilistic predictions, enabling users to assess the confidence of the model's classifications.

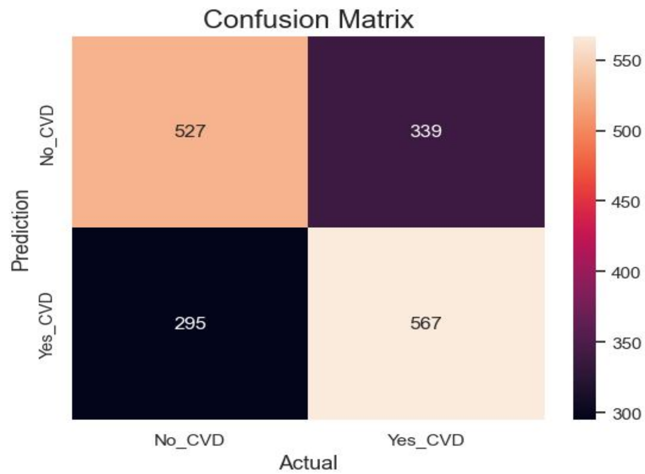

Figure 13. Confusion Matrix for Logistic Regression Algorithm

6) *Gradient Descent Algorithm:* Gradient descent is a fundamental optimization algorithm widely used in machine learning and optimization tasks. Its primary objective is to minimize a cost or loss function iteratively by adjusting the model parameters. The algorithm's name is derived from its strategy of descending along the steepest direction of the gradient of the cost function. In essence, it aims to find the minimum of a function by repeatedly moving towards the negative gradient. The process begins with initializing the model parameters randomly. At each iteration, the algorithm computes the gradient of the cost function for these parameters. The gradient indicates the direction of the steepest ascent, so the algorithm adjusts the parameters in the opposite direction to minimize the cost. The size of each parameter update is controlled by the learning rate, a hyperparameter that influences the step size in the parameter space. One of the key strengths of gradient descent lies in its applicability to a wide range of optimization problems, including linear and non-linear models. There are variations of gradient descent, such as stochastic gradient descent (SGD) and mini-batch gradient descent[5], which use subsets of the training data to update parameters, making it computationally more efficient, especially in large datasets. However, the choice of a suitable learning rate is crucial, as a too-small rate may lead to slow convergence, while a too-large rate can cause oscillations or divergence. Additionally, gradient descent may converge to local minima, and addressing this challenge often involves using more advanced optimization techniques or carefully selecting initialization strategies. Despite these considerations, gradient descent remains a foundational optimization algorithm, forming the basis for many sophisticated optimization methods used in machine learning.

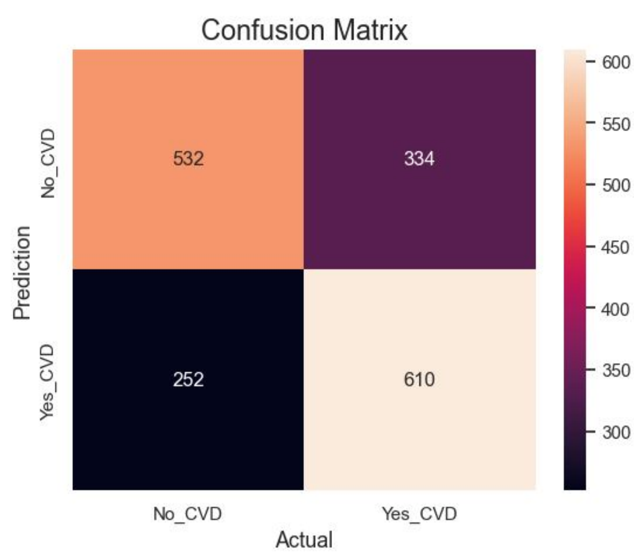

Figure 14. Confusion Matrix for Gradient Decent Algorithm

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

7) *XGBoost Algorithm:* XGBoost, short for eXtreme Gradient Boosting, is a powerful and widely used machine learning algorithm known for its efficiency, flexibility, and exceptional performance across various tasks. It belongs to the ensemble learning category, specifically boosting, where weak learners (usually decision trees) are sequentially trained to correct the errors of their predecessors[11]. XGBoost has gained popularity in both structured and tabular data scenarios and has won numerous machine-learning competitions due to its robustness. One key feature of XGBoost is its ability to handle complex relationships within data while mitigating overfitting. It achieves this by incorporating regularization terms in the objective function, controlling the complexity of individual trees and penalizing large coefficients. This regularization, combined with the algorithm's capacity for parallel processing, makes XGBoost computationally efficient and scalable, capable of handling large datasets XGBoost's strength lies in its versatility, as it supports both regression and classification tasks. It allows users to define custom optimization objectives and evaluation criteria, providing flexibility for a wide range of applications. Additionally, it includes native support for missing data, enabling the model to effectively handle incomplete datasets. The algorithm's success can be attributed to its implementation of gradient boosting principles, with enhancements such as tree pruning, handling missing values, and a robust method for selecting split points. XGBoost also introduces a novel regularization term called "shrinkage" or "learning rate," which controls the contribution of each tree to the final prediction. This feature aids in preventing overfitting and improving the model's generalization capabilities[10]. In summary, XGBoost stands out as a reliable and high-performance algorithm in the machine learning landscape, frequently chosen for its efficiency, scalability, and ability to deliver state-of-the-art results across various domains.

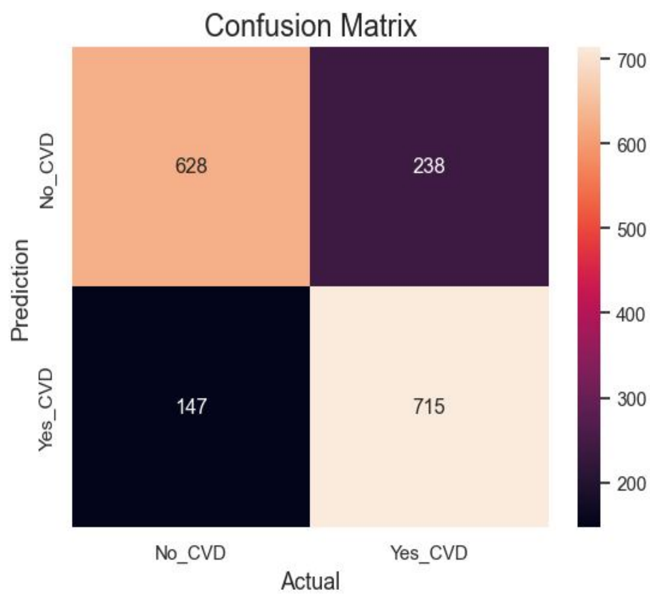

Figure 15. Confusion Matrix for XGB Algorithm

H. ROC-AUC Curve

The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) is a graphical representation commonly used in the evaluation of machine learning models, particularly those involved in binary classification tasks such as predicting the presence or absence of a medical condition like heart disease. The ROC-AUC curve provides a visual summary of a model's performance by illustrating the trade-off between its sensitivity and specificity across various decision thresholds. The ROC-AUC curve is generated by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at different classification thresholds. A higher AUC value, ranging from 0 to 1, indicates superior model performance. An AUC of 0.5 represents a model performing at chance, while an AUC of 1.0 reflects perfect discrimination between positive and negative instances.

Interpreting the ROC-AUC curve involves assessing the model's ability to discriminate between positive and negative cases. A curve that hugs the top-left corner signifies excellent discrimination, while a curve that closely follows the diagonal line suggests poor performance. Researchers and healthcare professionals can use the ROC-AUC analysis to choose an optimal threshold for the model, balancing sensitivity and specificity based on the specific goals and constraints of the healthcare application. Overall, the ROC-AUC curve serves as a valuable tool for understanding and communicating the performance characteristics of machine learning models in the context of cardiovascular health prediction and intervention**.**

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Below graphical representation of ROC-AUC Curve for different algorithms.

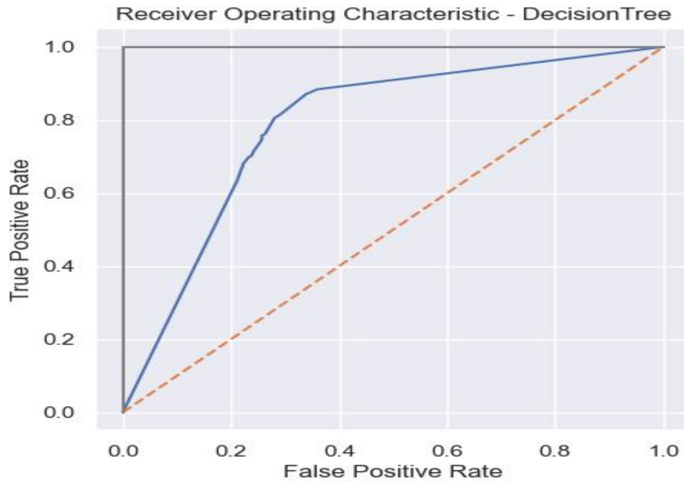

Figure 16. ROC-AUC Curve for Decision Tree

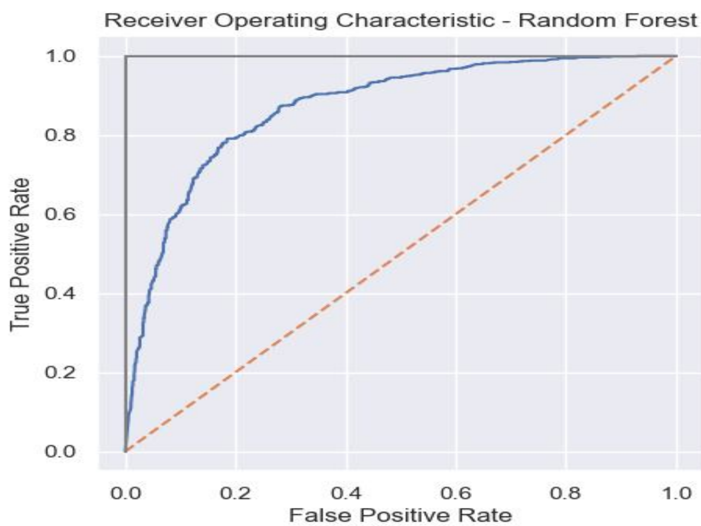

Figure 17. ROC-AUC Curve for Random Forest

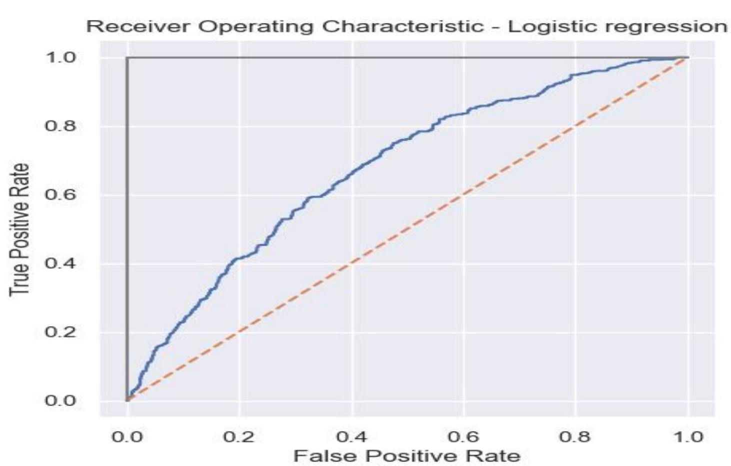

Figure 18. ROC-AUC Curve for Logistic Regression

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

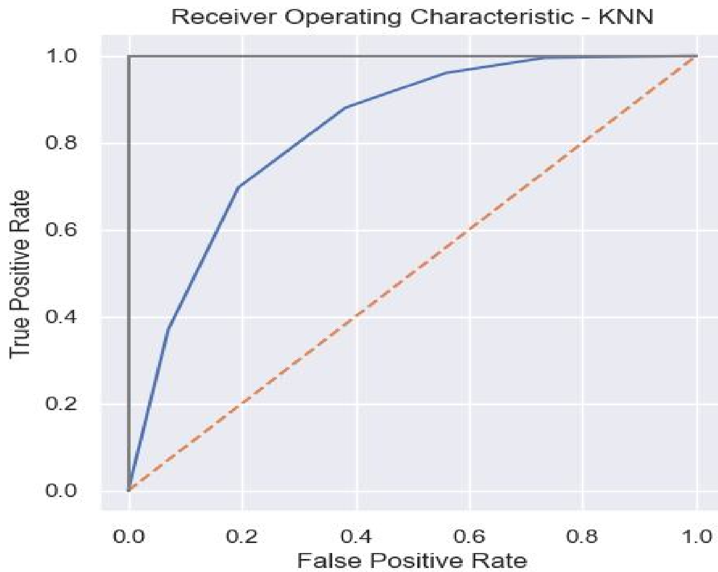

Figure 19. ROC-AUC Curve for K Nearest Neighbours

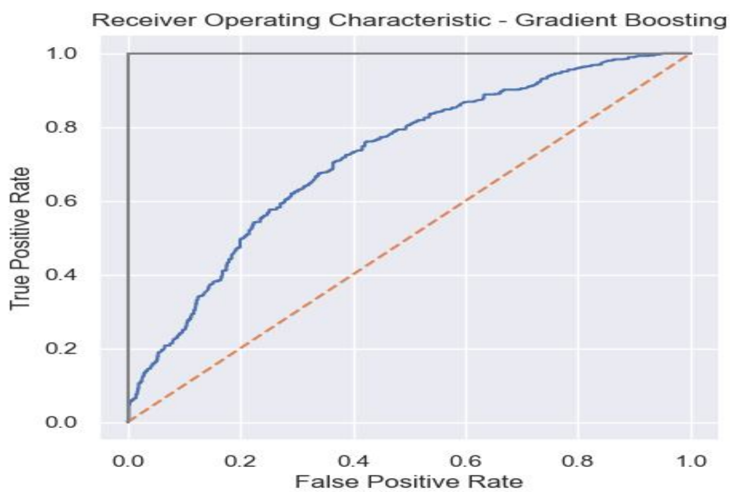

Figure 20. ROC-AUC Curve for Gradient Boosting

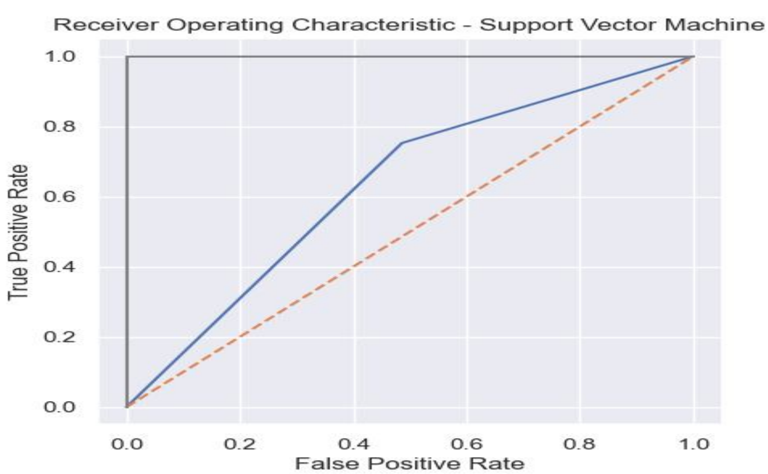

Figure 21. ROC-AUC Curve for Support Vector Machine

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

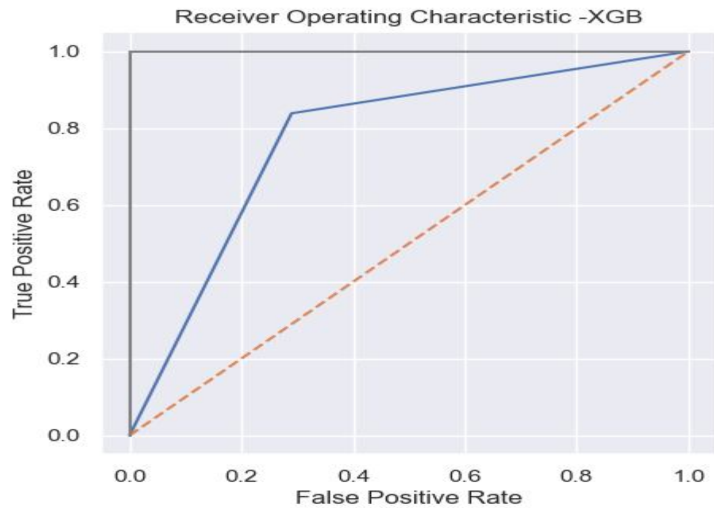

Figure 22. ROC-AUC Curve for XGB

IV. RESULT

In Table 1, After comparing all the algorithms with training accuracy, testing accuracy, and ROC-AUC Score. We find that the Random Forest Model is our best building model.

	Model_Name	Training_Accuracy(%)	Testing_Accuracy(%)	ROC_AUC_Score(%)
0	Random Forest	94.34	78.30	86.55
1	Decision Tree	94.34	76.62	79.56
2	XG Boost	88.36	76.44	76.45
3	KNN	84.37	74.19	81.84
4	Gradient Booster	69.28	67.41	73.10
5	Logistic Regression	63.67	63.54	68.36
6	Support Vector Machine	63.64	63.77	63.80

Table 1. Compared all Algorithms with Accuracies and ROC-AUC Curve

V. CONCLUSION

In conclusion, the application of a machine learning approach for predictive analysis and personalized intervention in revolutionizing cardiovascular health holds immense promise and potential. The utilization of advanced data-driven techniques has enabled a paradigm shift in our understanding and management of heart disease. The predictive models developed through machine learning not only offer enhanced accuracy in assessing cardiovascular risk but also pave the way for personalized interventions tailored to individual patient profiles.

The integration of artificial intelligence in cardiovascular health signifies a move towards more proactive and personalized healthcare strategies. By leveraging predictive analytics, healthcare professionals can identify high-risk individuals at an earlier stage, facilitating timely interventions and preventive measures. This proactive approach has the potential to significantly reduce the burden of heart disease, improve patient outcomes, and contribute to the overall efficiency of healthcare systems.

However, it is crucial to acknowledge the challenges and ethical considerations associated with this transformative approach. Addressing biases in machine learning models, ensuring the privacy and security of patient data, and maintaining transparency in decision-making processes are essential components of responsible implementation.

As we move forward, continued research and development in this field will be essential for refining and optimizing machine learning models, expanding the range of data sources, and validating the long-term effectiveness of personalized interventions. The fusion of cutting-edge technology with healthcare practices exemplifies a new era in cardiovascular health management, offering a more precise and patient-centric approach to combating heart disease and improving the overall well-being of individuals worldwide.

REFERENCES

[1] Cardiovascular diseases (CVDs). http://www.who.int/newsroom/factsheets/detail/cardiovascular-diseases-(cvds accessed on 30/9/2018. [Google Scholar].

[2] Patel B, Sengupta P. Machine learning for predicting cardiac events: what does the future hold? Expert Rev Cardiovasc Ther. 2020;18(2):77–84.

[3] Ghosh, P, Azam, S, Jonkman, M, Karim, A, Shamrat, FJM, Ignatious, E, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access. (2021) 9:19304–26. doi: 10.1109/ACCESS.2021.305375.

[4] Beunza J-J, Puertas E, García-Ovejero E, Villalba G, Condes E, Koleva G, et al. Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). J Biomed Inform. 2019;97:103257. [

[5] Shah D, Patel S, Bharti SK. Heart Disease Prediction using Machine Learning Techniques. SN Computer Sci. 2020;1:345–6.

[6] Al Aref, SJ, Anchouche, K, Singh, G, Slomka, PJ, Kolli, KK, Kumar, A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J. (2019) 40:1975–86. doi: 10.1093/eurheartj/ehy404.

[7] Krittanawong, C, Virk, HUH, Bangalore, S, Wang, Z, Johnson, KW, Pinotti, R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. (2020) 10:1–11. doi: 10.1038/s41598-020-72685-1.

[8] Mezzatesta, S, Torino, C, De Meo, P, Fiumara, G, and Vilasi, A. A machine learning-based approach for predicting the outbreak of cardiovascular disease in patients on dialysis. Comput Methods Prog Biomed. (2019) 177:9–15. doi: 10.1016/j.cmpb.2019.05.005

[9] Kaur P, Kumar R, Kumar M. A healthcare monitoring system using random forest and internet of things (IoT). Multimed Tools Appl. 2019;78:19905–16.

[10] Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. 2019;16:100203.

[11] Pal M, Parija S. Prediction of Heart Diseases using Random Forest. J Physics: Conf Ser. 2021;1817:012009. 10.1088/1742-6596/1817/1/012009.

[12] Allan, S, Olaiya, R, and Burhan, R. Reviewing the use and quality of machine learning in developing clinical prediction models for cardiovascular disease. Postgrad Med J. (2022) 98:551–8. doi: 10.1136/postgradmedj-2020-139352

[13] Maini, E., Venkateswarlu, B., and Gupta, A. (2018). "Applying machine learning algorithms to develop a universal cardiovascular disease prediction system" in International Conference on Intelligent Data Communication Technologies and Internet of Things. Springer, Cham. 627–632.

[14] Smita, and Kumar, E. Probabilistic decision support system using machine learning techniques: a case study of cardiovascular diseases. J Discret Math Sci Cryptogr. (2021) 24:1487–96. doi: 10.1080/09720529.2021.1947452

[15] Shu, S, Ren, J, and Song, J. Clinical application of machine learning-based artificial intelligence in the diagnosis, prediction, and classification of cardiovascular diseases. Circ J. (2021) 85:1416–25. doi: 10.1253/circj.CJ-20-1121.