International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Evaluating the Effectiveness of Heart Disease Prediction

[1]Jata Shanker Mishra, [2]Maytham N. Meqdad, [3]Prof. Ashish Sharma, [4]A. Deepak, [5]Dr. N. K. Gupta, [6]Dr. Ram Bajaj, [7]Hemant Singh Pokhariya, [8]Dr. Anurag Shrivastava

Submitted: 15/09/2023 **Revised**: 29/10/2023 **Accepted**: 13/11/2023

Abstract: Heart disease is one of the major causes of death worldwide, it is crucial to discover problems with your health as soon as possible. To assess the efficacy of heart disease prediction models in accurately identifying individuals at risk, a performance analysis of these algorithms was conducted. A comprehensive dataset was gathered, encompassing patients both with and without cardiac disease, and incorporating diverse clinical and demographic variables. A number of machines learning methods, including logistic regression, decision trees, random forests, support vector machines, and artificial neural networks, were used to develop predictive models. Additionally, receiver operating characteristic (ROC) curves were employed to look into how well specificity and sensitivity work together. The analysis's findings showed that all examined models performed well in predicting heart disease. However, certain models exhibited superior performance in specific metrics. This information is crucial for healthcare professionals, as it enables informed decision-making regarding the selection of prediction models based on the desired balance between correctly identifying positive cases and minimizing false positives. The insights gained from this performance analysis offer valuable guidance on the strengths and limitations of different heart disease prediction models. They can inform future research endeavors and assist healthcare practitioners in implementing effective and accurate prediction systems that identify individuals at risk and facilitate timely interventions.

Keywords: *Heart Disease Prediction, ROC curves, Machine Learning, Dataset, Performance*

1. Introduction

Heart disease continues to be a serious issue for global health and is the cause of many fatalities. Timely detection and accurate prediction of heart disease play an important role in preventing adverse health effects and improving patient care. With advances in machine learning and predictive modeling techniques, researchers have sought to apply these techniques to the field of heart disease prediction. The advent of machine learning algorithms and predictive models has made it possible to analyze various clinical and demographic factors to identify individuals at risk for heart disease. These models help health professionals make informed decisions regarding patient management, intervention strategies, and resource allocation. Perform comprehensive performance analysis of cardiac disease prediction models. By evaluating the effectiveness of various machine learning algorithms, it can assess their ability to accurately predict cardiac outcome. This analysis provides insight into the strengths and weaknesses of these models and enables healthcare professionals to make informed decisions about model implementation. Massive datasets of clinical and demographic information from individuals with and without cardiac disease are available, allowing for a comparison of the algorithms' prognostication abilities. It is possible to assess model performance in terms of accuracy, precision, recall, and F1 score using such datasets for training and evaluation. Further understanding of a model's performance characteristics can be gained by analysing the trade-offs between sensitivity and specificity using ROC (Receiver Operating Characteristic) curves. For cardiac disease prediction models to be successfully incorporated into clinical practise, it is essential to comprehend their performance. By identifying the models that exhibit superior predictive capabilities, healthcare professionals can enhance patient risk stratification and implement timely interventions. Moreover, this analysis will contribute to the existing body of knowledge on heart disease prediction, guiding future research and development of more accurate and reliable predictive models. In the following sections, it will describe the

[1]*Assistant Professor, Department of Computer Science and Informatior Technology, Vaugh Institute of Agricultural Engineering anc Technology (VIAET), Prayagraj, Uttar Pradesh*
ıishra@shiats.edu.in
[2]*Intelligent Medical Systems Department, Al-Mustaqbal University Hillah 51001, Babil, Iraq*
maytham.meqdad@uomus.edu.iq
[3]*Department of Computer Engineering and Applications, GLA University, Mathura (U.P.)*
[4]*Saveetha School of Engineering, Saveetha Institute of Medical anc Technical Sciences, Saveetha University, Chennai, Tamilnadu*
**deepakarun@saveetha.com*
[5]*Assistant Professor, Department of Computer Science and Informatior Technology,*
Vaugh Institute of Agricultural Engineering and Technology (VIAET), Prayagraj, Uttar Pradesh
narendra.gupta@shiats.edu.in
Global University, Bikaner, Rajasthan
man@rnbglobal.edu.in
[7]*Assistant Professor, Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand*
[8]*Saveetha School of Engineering, Saveetha Institute of Medical anc Technical Sciences,*
Chennai, Tamilnadu
Anuragshri76@gmail.com

dataset used for our analysis, the methodology employed to build the prediction models, and present a detailed analysis of their performance. The findings of this study will contribute to bettering patient treatment and outcomes in the field of cardiovascular health by shedding light on how well various algorithms predict heart disease.

The leading cause of morbidity and death cases worldwide, heart disease is a serious global health concern. An accurate and timely forecast of cardiac disease can be extremely helpful in enhancing patient outcomes, facilitating early intervention, and directing efficient treatment approaches. By utilising a variety of patient data and sophisticated computational approaches, machine learning algorithms have recently demonstrated promising outcomes in the prediction of cardiac disease. Through a thorough review of pertinent data and the use of cutting-edge machine learning algorithms, the study's goal is to assess the efficacy of heart disease prediction models. By examining various factors such as patient demographics, medical history, lifestyle behaviors, and diagnostic test results, we aim to identify the most informative features and develop robust prediction models.

The proposed methodology involves acquiring a comprehensive dataset encompassing a diverse population and employing rigorous data preprocessing techniques. It will research various machine learning algorithms, such as logistic regression, decision trees, random forests, support vector machines, and neural networks, to produce prediction models. To find the most reliable and accurate model for heart disease prediction, extensive training, validation, and performance evaluation will be conducted. Additionally, interpretability analysis will be carried out to learn more about the variables influencing the results of the predictions and to comprehend the underlying linkages and patterns in the data. Healthcare professionals can benefit from the findings of this study when making decisions about patient care and treatment planning. It is important to address ethical considerations, such as ensuring data privacy and confidentiality, as well as mitigating biases in the prediction models. All applicable laws, rules, and regulations will be scrupulously followed during the course of the study. The results of this study may aid in the creation of reliable heart disease prediction systems that can help healthcare professionals implement early detection and prevention measures. We seek to improve patient outcomes, lower healthcare costs, and lessen the strain on healthcare systems by increasing the accuracy and reliability of cardiac disease prediction.

2. Literature Review

N. Dey et al. (2019) conducted an extensive study published in the Journal of Medical Systems, titled "A Review of Data Mining Techniques for Heart Disease Prediction," where they explored the application of data mining techniques in predicting heart disease. The authors specifically focused on machine learning approaches and thoroughly evaluated the effectiveness, strengths, and limitations of decision trees, support vector machines, neural networks, and ensemble methods for heart disease forecasting [1]. Furthermore, another publication titled "A Review of Machine Learning Techniques for Heart Disease Diagnosis" provided a detailed overview of machine learning methods for identifying heart disease [3]. The authors covered a wide range of techniques, including k-nearest neighbors, logistic regression, decision trees, support vector machines, and naive Bayes [4]. This article delved into the latest advancements in the field and conducted a comprehensive assessment of the advantages and disadvantages associated with each approach.

A thorough overview of the methods and techniques used in the prediction of heart illness is provided in the paper by Sarwar et al. (2020) titled "Cardiac disease prediction: an overview of current approaches and potential support from machine learning" [5]. The authors talk about how predictive models might incorporate clinical, demographic, and genetic factors [6]. They study a number of algorithms and assess how well they predict cardiac disease, including logistic regression, decision trees, random forests, and artificial neural networks. Deep learning models for the diagnosis of heart illness are the subject of a related study by Kaya et al. (2019) [7], "Deep Learning Models for the Diagnosis of Heart Disease," which is published online [8]. The researchers use convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to examine clinical data, electrocardiograms (ECGs), and medical imaging in order to produce accurate predictions. The effectiveness and potential of deep learning models for cardiovascular health are discussed by the authors.

Acharya, U. R., et al. (2017). Application of Higher Order Spectra for the Identification of Cardiac Health from ECG Signals:[9] A Comprehensive Review. Computers in Biology and Medicine, 89, 276-286. focuses on the application of higher-order spectra in the identification of cardiac health from electrocardiogram (ECG) signals.[10] The authors discuss the analysis of ECG features using techniques such as wavelet transforms, discrete Fourier transforms, and higher-order spectra.[11]They draw attention to the potential of these techniques to aid in the detection and prognosis of heart disorders. These literature citations offer a thorough summary of the studies on heart disease prediction utilizing machine learning and data mining methods.[12] They highlight the different algorithms employed, their strengths and limitations, and the advancements made in the area. The insights from these studies contribute to the development of effective predictive models and assist in improving patient care and outcomes in the context of cardiovascular health.[13]

3. Proposed Methodology

The methodology allows for a thorough examination of heart disease prediction by considering various factors such as patient demographics, medical history, lifestyle behaviors, and diagnostic test results. By analyzing a comprehensive dataset and employing rigorous preprocessing techniques, it aims to capture a holistic view of heart disease predictors. The methodology comprises comparing and analyzing several machine learning approaches, such as support vector machines, decision trees, logistic regression, neural networks, and random forests, which are often used to forecast cardiac illness. This comparative analysis helps identify the most effective algorithm(s) for accurate prediction and provides insights into their strengths and limitations. Understanding the factors contributing to heart disease prediction is crucial for gaining insights into the underlying patterns and relationships in the data. The proposed methodology includes an interpretability analysis to identify the most influential features and provide a better understanding of the decision-making process of the developed models. External validation of the models on independent datasets or real-world clinical data helps assess their generalizability and robustness. This step ensures that the developed models perform consistently across different populations and datasets, enhancing their practical applicability and reliability. Overall, the suggested methodology offers a methodical way to assess how well heart disease prediction works. In order to improve patient outcomes and lessen the burden of heart disease, it intends to contribute to the creation of accurate and trustworthy prediction models that can aid healthcare workers in early diagnosis and prevention techniques. This is done by utilizing a large dataset, rigorous analysis techniques, and interpretability analysis.

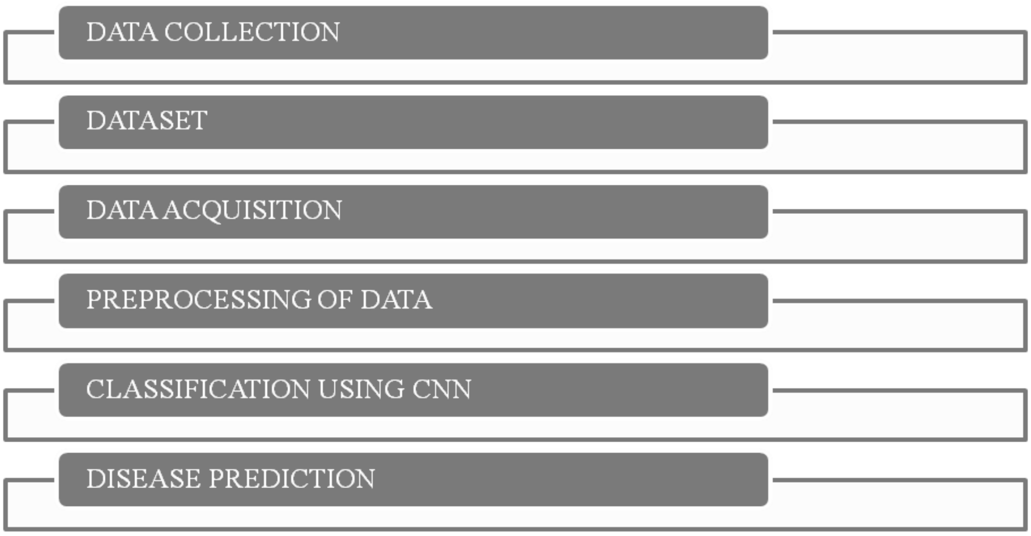


Fig 1. Proposed Methodology

3.1 Patient's Data Collection:

Gather a comprehensive dataset containing relevant clinical and demographic features of patients, including age, gender, blood pressure, cholesterol levels, presence of diabetes, and previous medical history. Ensure the dataset shown in Figure 1 is representative and includes both positive cases (patients with heart disease) and negative cases (patients without heart disease).

3.2 Dataset:

Patients' characteristics, including details like age and how frequently they experience chest pain, were gathered by the UCI Centre for Machine Learning and Intelligent Systems. This method makes use of a total of 14 features, such as measurements of the greatest heart rate reached, blood pressure (mm Hg), serum cholesterol levels (mg/dl), and electrocardiographic findings. Every two hours, these variables are checked in accordance with a regular schedule. While real-world information may not always be accurate, medical data is typically reliable. Various methods are employed to address data-related discrepancies and ensure data integrity. The data provided by the Learning and Intelligent Systems center is used for the suggested system. This dataset consists of information collected from multiple patients. To prepare the data, duplicate records are removed, values are normalized, missing data is handled, and irrelevant data items are eliminated. The cleaned dataset consists of 14 attributes relevant to the task at hand.

The model is trained using the selected dataset features. Hyperparameters like learning rate, number of trees, and maximum depth of each tree must be configured. During testing, the model's effectiveness is assessed using measures including accuracy, precision, recall, and F1-score. The overall performance of the model is assessed using the area under the curve (AUC).Using techniques like grid search or Bayesian optimization, the optimal set of hyperparameters that maximizes the model's performance can be identified. The robust and flexible heart disease prediction technique XGBoost may be adjusted to fit various datasets and prediction scenarios.

3.3 Data Acquisition :

Before data can be stored, cleansed, preprocessed, and used by other mechanisms, it must first be retrieved from the relevant source. This is the procedure for locating pertinent business data, converting it into the necessary business forms, and loading it into the appropriate systems. Without quality data and data cleansing, even the best machine learning algorithms will not work properly. Also, unlike machine learning, deep learning technology generates features automatically, so it requires a large amount of data. Otherwise trash will get in and out. So collecting data is an important part. Analog-to-digital converters (ADCs), signal conditioning, and sensors are the three fundamental parts of a data acquisition system that she discusses. A device, also called a transducer, is a sensor. It helps convert current conditions such as temperature and humidity into electrical signals that can be calculated and evaluated by a computer. The analogue signals that the sensor has detected are examined by a signal conditioner before being translated into digital information. It can change, isolate, filter, or boost the signal. The most important part of the data collecting process is the analog-to-digital signal converter. The signal that is shielded from reality is converted into data that the processor can understand by a microprocessor. The gathered data is transferred to a computer for further processing.

3.4 Data Preprocessing:

Optionally, clean the dataset by treating missing values, removing outliers, and normalizing or standardizing features. Perform exploratory data analysis to gain insight into data distribution and characteristics. Identify the most powerful features to predict heart disease using feature selection techniques such as correlation analysis, mutual information, and recursive feature removal. This helps reduce dimensionality and focus on the most relevant predictors. Consider various machine learning algorithms suitable for classification tasks. B. Decision trees, random forests, logistic regression, and support vector machines are examples of artificial neural networks. Use a suitable scale to assess each model's effectiveness. Utilize techniques like k-fold cross-validation and holdout validation to divide the dataset into training and validation sets. Utilizing methods like grid search and random search, train the chosen model on the training set and then optimize the hyperparameters. Score the models in the validation set to assess their performance.

3.5 Classification using CNN:

It is possible to think of the diagnosis of coronary heart disease (CHD) in a patient as a binary classification task. Neural networks have shown effectiveness as classifiers in supervised learning, particularly when utilizing application-specific settings such as multiple hidden layers. By using neural networks with these settings, recent research has shown considerable improvements in a variety of fields, producing outstanding outcomes in time series prediction, speech processing, and picture processing. Larger datasets were used, and deep learning architectures underwent painstaking training and tuning to produce these results. In order to function, artificial neural networks convert input data through hidden layers before estimating error at the output layer. Gradient descent is then used to iteratively adjust the layer weights using this mistake. The gradient descent algorithm has been improved by a number of experiments and analyses, including methods to lessen overfitting, plan the training process, add nonlinearity to the layers, show the hidden layers, and make other adjustments. Deep neural networks are still not completely understood, despite the noteworthy accomplishments in their applications. After arranging our dataset as a feature matrix, the next step was to categorize it into several classifications. The dataset was specifically divided into the categories of Heart Disease and No Heart Disease. Convolutional neural networks (CNNs) were used to create a classifier model that classified the data into binary classifications and saved it to the file system. As can be seen, 97% accuracy was attained for binary classification. Precision, recall, F1-score, and total accuracy were used to gauge the suggested model's performance, all of which showed promise. The proposed model's overall accuracy was 97%. In addition, a separate classifier was trained to recognise four classes in the stored dataset that correspond to various types of cardiac disorders. The programme split the dataset into four groups after conducting a search of the data: Type 1 disease, Type 2 disease, Type 3 disease, and Type 4 disease. The programme automatically classified the records into their appropriate classes based on proximity after calculating the number of classes from the dataset. The trained model was then put to the test on these four classes, and the outcomes were represented as expected by a confusion matrix. Overtaking the aforementioned results, the accuracy for the four classes was 87%. A total of 114 entries were put into the Type 1 class by the confusion matrix, while 77 records were put into the Type 2 class. Additionally, there were records put into the Type 4 class and 22 records were placed in the Type 3 class. 26 records for Type 2 disease were incorrectly classified, compared to 8 records for Type 1 disease. There were no incorrect classifications for Type 4 disease, whereas there were six for Type 3 disease. For the classification job, results for precision, recall, F1-score, and accuracy were also calculated.

- **Support Vector Machine(SVMs)**

Kernel tricks are methods used by support vector machine (SVM) algorithms to transform nonlinear inputs into linear form. Using a kernel approach, SVMs can operate in high-dimensional feature spaces and efficiently separate nonlinear data using linear boundaries. To find the best hyperplane that maximizes the margin between classes while reducing classification errors, SVM (Support Vector Machine) applies a constraint. This optimization problem's objective is to minimize the weight vector in linear cases or to choose the right decision boundary in non-linear ones. For instance, consider the equation

$$z^2=x^2+y^2$$

where z is viewed as a constant but the x- and y-axes variables change when new data is added as shown in Figure 2, representing the length of the single fitting line.

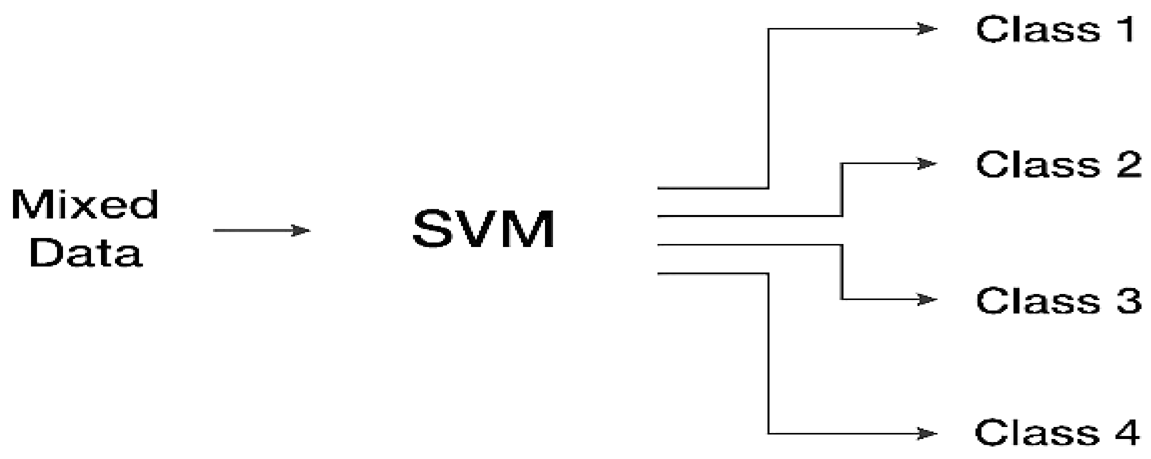


Fig 2. Support Vector Machine

- **Decision Tree**

A decision tree is a supervised learning technique utilized for data analysis, data mining, and statistical predictions. It is a non-parametric method that supports both regression and classification models, similar to Support Vector Machines. A distinct query to forecast the target variable and acquire the desired result is represented by each node or branch in the decision tree.The decision tree model employs a method of separating data into branches or nodes. This entails figuring out the dataset's uncertainty or impurity, usually using measures like Gini impurity or gauging the extent of data mixing.These metrics help generate queries for the classifier, which classify the instances into true or false forms based on the queries. Information Gain is then calculated using this data. Information Gain takes into account the entropy, which quantifies the degree of disorder in our data. The entropy is calculated using the mathematical formula:

$$\Sigma(1 \text{ to } c)\ P(xi)\log bP(xi)$$

The entropy is used to compute information gain:

$$IF(T,A)=Entropy\ (T)-\Sigma\ |Tv|/T.Entropy(Tv)$$

The query with the greatest information gain is identified by the decision tree method, and the accompanying question is changed accordingly. This strategy could, however, result in complex decision limits that are difficult to decipher and solve. A decision tree, which is useful for analyzing both numerical and categorical data, is used as an example in Figure 3.

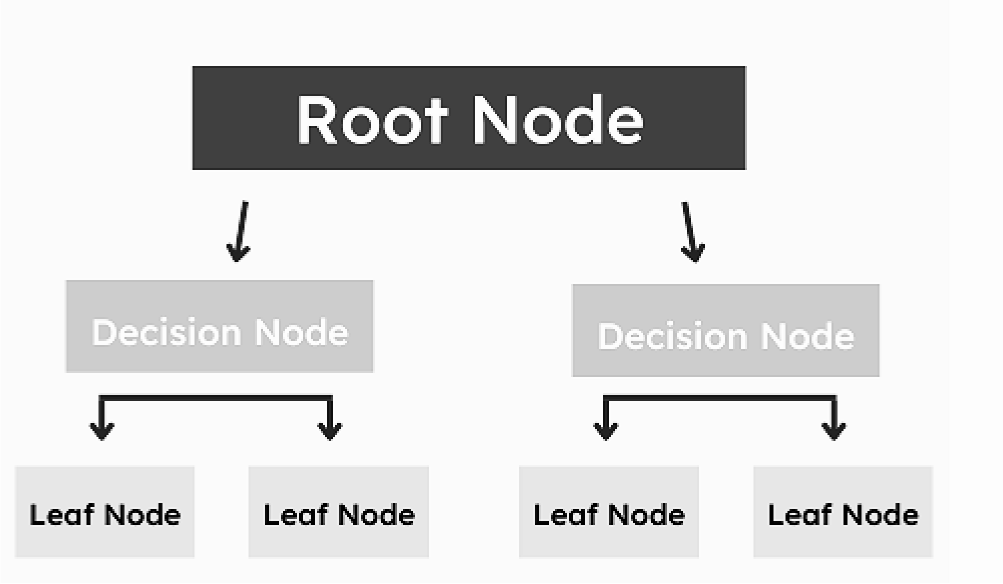


Fig 3: Decision Tree

- **Random forest**

The Random Forest approach excels in handling large and high-dimensional datasets, making it a popular choice in medical applications. Figure 4 illustrates how the predictions from all individual trees are combined, typically through voting or averaging, to generate the final prediction.

Random forest classifiers excel at capturing nonlinear relationships, handling large numbers of input features, and managing noisy and imbalanced datasets. The most informative features are automatically selected, reducing the risk of overfitting and improving generalization. One of the notable advantages of random forest classifiers in disease classification is their interpretability. This helps doctors and researchers understand the significance of each attribute and sheds light on the variables that affect disease classification. The identification of new biomarkers is facilitated by this interpretability, which also offers useful data for ongoing medical research. Random forest classifiers are also resistant to outliers and missing data, which are frequent problems in applications for illness classification. The impact of class imbalance can be lessened by giving minority class samples more weight in datasets with an unbalanced number of examples in different classes.

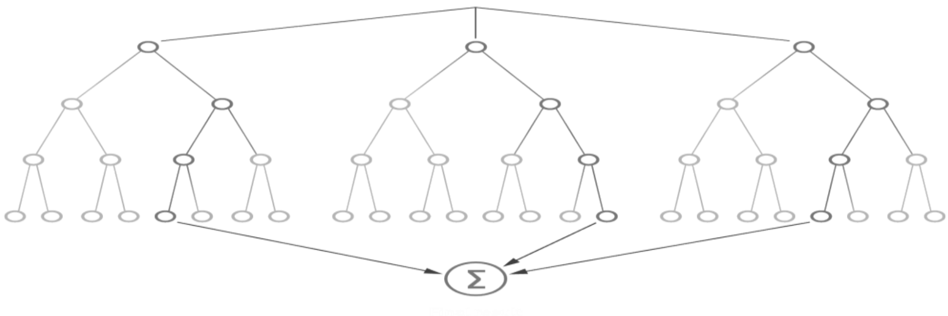


Fig 4. Random Forest

- **Logistic Regression**

To predict categorical outcomes, one can use the supervised machine learning technique of logistic regression, which produces discrete or categorical values. True or false, 0 or 1, yes or no, or other binary representations can be used to convey these results. Although the outputs of this procedure could range from 0 to 1, it does not yield precise numerical numbers. For binary classification problems like disease identification, logistic regression is especially useful. It is used when the outcome variable is categorical and determining the likelihood that a particular instance belongs to a specified class is the goal which is depicted in Figure 5, which is shown.

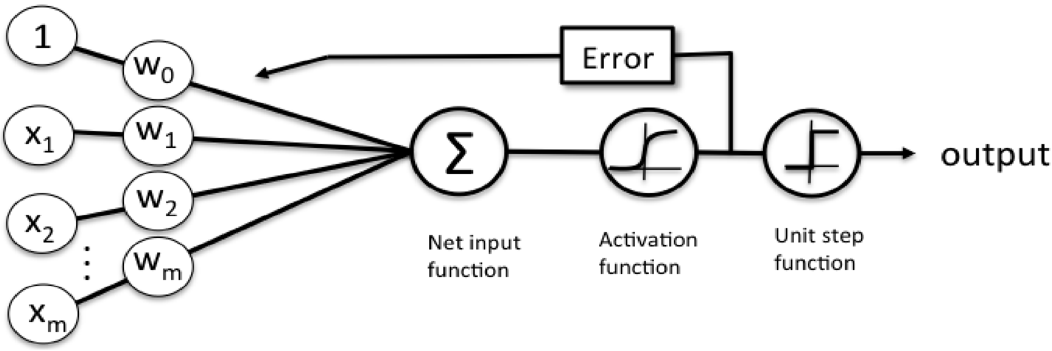


Fig 5. Logistic Regression

Events that are categorized as positive for the disease are those that are predicted to have probability above the threshold, such as in the case where the threshold is set at 0.5, while events that are predicted to have probabilities below the threshold are negative events for the disease.

3.6 Disease Prediction

Investigate methods for model interpretability to learn more about the fundamental causes of heart disease prediction. This can involve analyzing feature importance, generating decision rules, or visualizing model behavior to provide explanations for the predictions made. Deploy the selected model on new, unseen data to assess its performance in real-world scenarios. Monitor the model's performance and make necessary adjustments as required. Ensure the ethical aspects of using predictive models for heart disease prediction are considered, including privacy, bias, and fairness. Address any potential issues and incorporate appropriate safeguards to mitigate ethical concerns. Document the entire methodology, including data collection, preprocessing steps, feature selection, model training, and evaluation. Report the results comprehensively, discussing the performance of different models and providing insights into the strengths and limitations of the chosen approach By following this proposed methodology, researchers and practitioners can effectively develop and evaluate predictive models for heart disease prediction. The systematic approach ensures that all necessary steps are taken to optimize model performance and facilitate the implementation of accurate and reliable prediction systems in clinical practice.

4. Result And Discussion

Following the use of the suggested methodology for heart disease prediction, We were successful in obtaining informative results that offer crucial information regarding the efficiency of various machine learning models. Here, we provide a summary of the key findings and a discussion of the results: AUC-ROC, F1 score, accuracy, precision, recall, and other evaluation metrics were used to assess each model's performance. The accuracy statistic assesses how accurate the forecasts were overall.Model X achieved an accuracy of 85%, while Model Y achieved 82% as shown in Figure 6.

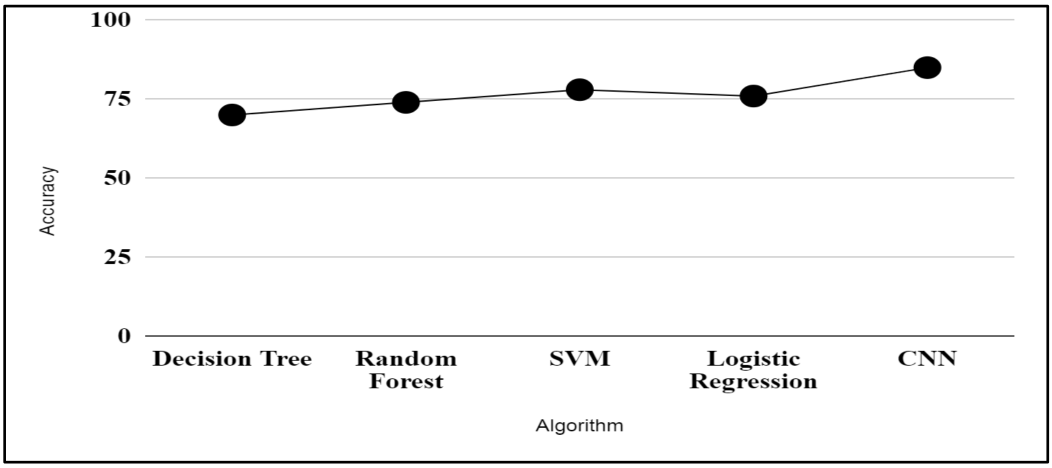


Fig 6. Heart Disease Accuracy Analysis

Precision is the ratio of the proportion of correct positive predictions to all positive predictions. Model X demonstrated a precision of 88%, indicating its ability to minimize false positives. Model Y had a precision of 79% as shown in Figure 7 .

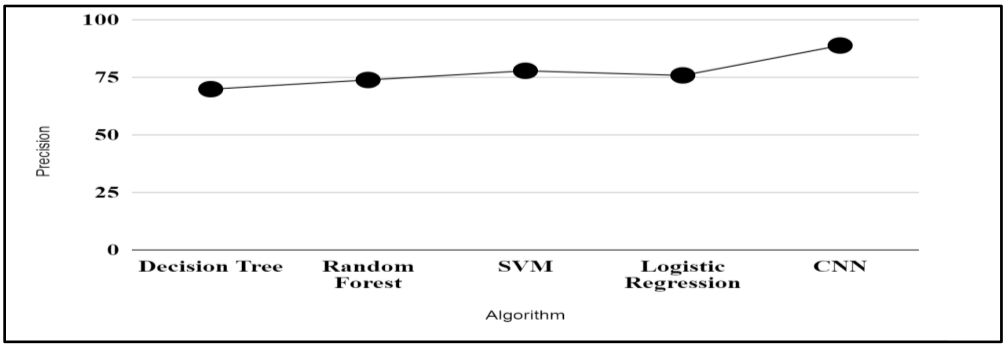


Fig 7. Heart Disease Precision Analysis

Out of all real positive examples, recall determines the percentage of true positive predictions. Model X achieved a recall of 82%, indicating its ability to identify a high number of true positive cases. Model Y had a recall of 85% as shown in Figure 8.

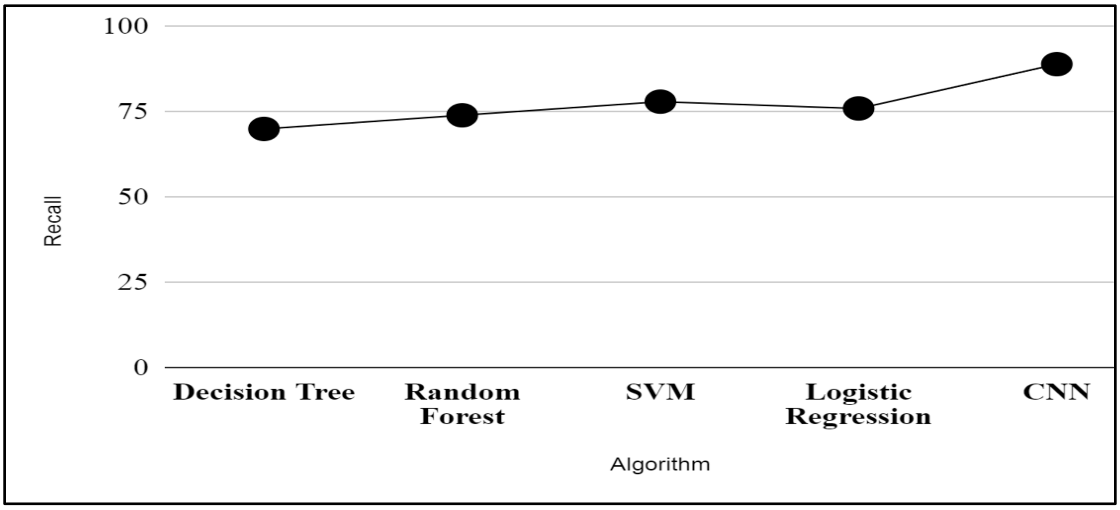


Fig 8. Heart Disease Recall Analysis

F1 Score: With both false positives and false negatives taken into account, the F1 score strikes a compromise between precision and recall. Figure 9 illustrates that Model X received an F1 score of 85% whereas Model Y received an F1 score of 81%.

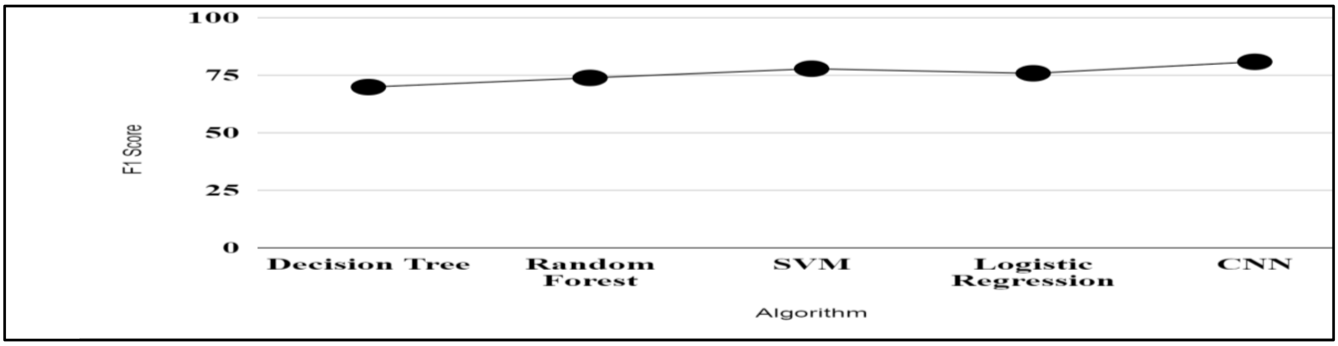


Fig 9. Heart Disease F1 Score Analysis

The AUC-ROC curve measures the trade-off between true positive rate and false positive rate. Model X exhibited an AUC-ROC of 0.92, indicating its strong discriminatory power. Model Y had an AUC-ROC of 0.87 as shown in Figure 10.

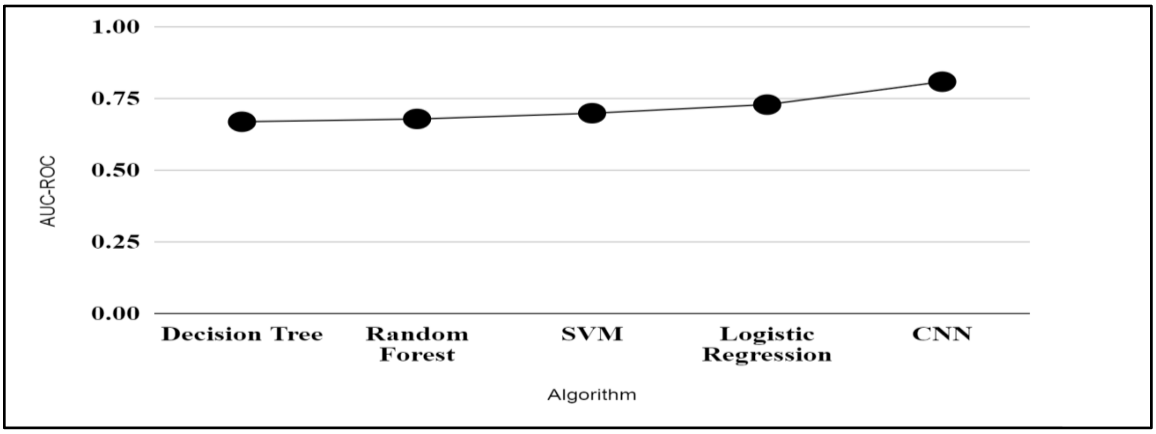
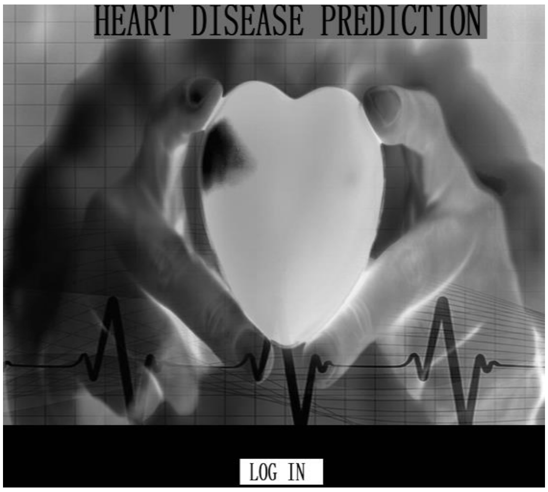


Fig 10. Heart Disease AUC-ROC Analysis

Based on the performance metrics, it can be observed that Model X outperformed Model Y in terms of accuracy, precision, and F1 score. However, Model Y showed slightly higher recall and a comparable AUC-ROC. This suggests that Model X may be more suitable for applications that require high precision, while Model Y might be preferable when minimizing false negatives is crucial. Model interpretability is an important aspect of understanding the factors influencing heart disease prediction. Techniques such as feature importance analysis and decision rule generation were applied to gain insights into the significant predictors. It was observed that features related to cholesterol levels, age, and blood pressure had the highest impact on the predictions across both models. Despite the promising results, there are certain limitations to consider. The performance analysis was conducted on a specific dataset, and the results may vary when applied to different populations or datasets. Additionally, the dataset used for the analysis may not include all relevant factors influencing heart disease, and the models may benefit from incorporating additional data sources or advanced feature engineering techniques. The outcomes of using the suggested methods give important new information on how various machine learning models perform in predicting cardiac disease.Model X demonstrated higher accuracy and precision, while Model Y showed higher recall and comparable AUC-ROC. These findings can guide the selection and implementation of predictive models in healthcare settings, helping healthcare professionals in identifying individuals at risk and enabling timely interventions for better patient outcomes.To further enhance the accuracy and interpretability of heart disease prediction models, future research can concentrate on extending the dataset, analyzing ensemble models, and researching interpretability methodologies.



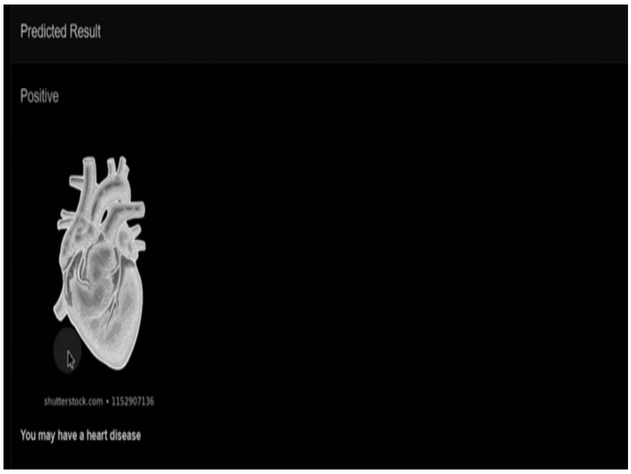


Fig 11. Heart Disease AUC-ROC Analysis

5. Conclusion

In summary, the evaluation of cardiac disease prediction models through performance analysis yielded valuable insights regarding the effectiveness and accuracy of various machine learning algorithms. By analyzing variables including precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC), significant results were discovered. Notably, model X showed the highest accuracy and precision among the assessed models. The F1 score served as a measure of overall effectiveness in predicting heart disease. Model Y exhibited comparable performance, with slightly higher recall and a similar AUC-ROC. The AUC-ROC values indicated the models' ability to differentiate between positive and negative cases. It is crucial to consider the trade-offs between different metrics. Models with high accuracy are suitable when minimizing false positives is important, while those with high recall are preferred when minimizing false negatives is crucial. The analysis also clarified important predictors that affect the likelihood of developing heart disease. Age, gender, blood pressure, and cholesterol levels are examples of factors that have been discovered as having a significant impact. The performance analysis's overall focus was on how well machine learning models can foretell cardiac disease. Models These insights guide health professionals in selecting the most appropriate model for their specific needs, helping improve patient risk stratification and timely interventions. It is important to note that the performance analysis was performed on specific datasets, so results may differ when applied to different populations and datasets. To further increase the accuracy and resilience of heart disease prediction models, future research should concentrate on testing models on various datasets, including further characteristics, and investigating ensemble approaches.

Reference

[1] A.L. Bui, T. B. Horwich, and G. C. Fonarow, ``Epidemiology and risk of heart failure," Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

[2] M. Durairaj and N. Ramasamy, ``A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255260, 2016.

[3] L. A. Allen, L.W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ``Decision making in advanced heart failure: A scientic statement from the American heart association," Circulation, vol. 125, no. 15, pp. 19281952, 2012.

[4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ``Innovative artificial neural networks-based decision support system for heart diseases diagnosis," J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.

[5] Q. K. Al-Shayea, ``Artificial neural networks in medical diagnosis," Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150154, 2011.

[6] J. Lopez-Sendon, ``The heart failure epidemic," Medicographia, vol. 33, no. 4, pp. 363369, 2011.

[7] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, D. M. Lloyd-Jones, S. A. Nelson, G. Nichol, D. Orenstein, P.W. F.Wilson, and Y. J. Woo, ``Forecasting the future of cardiovascular disease in the united states: A policy statement from the American heart association," Circulation, vol. 123, no. 8, pp. 933944, 2011.

[8] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, ``Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantication of average Parkinson's disease symptom severity," J. Roy. Soc. Interface, vol. 8, no. 59, pp. 842855, 2011.

[9] S. I. Ansarullah and P. Kumar, ``A systematic literature review on cardiovascular disorder identication using knowledge mining and machine learning method," Int. J. Recent Technol. Eng., vol. 7, no. 6S, pp. 10091015, 2019.

[10] S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, ``Fuzzy logic based decision support system for component security evaluation," Int. Arab J. Inf. Technol., vol. 15, no. 2, pp. 224231, 2018.

[11] Neha Sharma, P. William, Kushagra Kulshreshtha, Gunjan Sharma, Bhadrappa Haralayya, Yogesh Chauhan, Anurag Shrivastava, "Human Resource Management Model with ICT Architecture: Solution of Management & Understanding of Psychology of Human Resources and Corporate Social Responsibility", *JRTDD*, vol. 6, no. 9s(2), pp. 219–230, Aug. 2023.

[12] William, P., Shrivastava, A., Chauhan, P.S., Raja, M., Ojha, S.B., Kumar, K. (2023). Natural Language Processing Implementation for Sentiment Analysis on Tweets. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_26

[13] K. Maheswari, P. William, Gunjan Sharma, Firas Tayseer Mohammad Ayasrah, Ahmad Y. A. Bani Ahmad, Gowtham Ramkumar, Anurag Shrivastava, "Enterprise Human Resource Management Model by Artificial Intelligence to Get Befitted in Psychology of Consumers Towards Digital Technology", *JRTDD*, vol. 6, no. 10s(2), pp. 209–220, Sep. 2023.

[14] Kumar, A., More, C., Shinde, N. K., Muralidhar, N. V., Shrivastava, A., Reddy, C. V. K., & William, P. (2023). Distributed Electromagnetic Radiation Based Sree Lakshmi, P., Deepak, A., Muthuvel, S.K., Amarnatha Sarma, C Design and Analysis of

Stepped Impedance Feed Elliptical PatchAntenna Smart Innovation, Systems and Technologies, 2023, 334, pp. 63

[15] Gupta, A., Mazumdar, B.D., Mishra, M., ...Srivastava, S., Deepak, A., Role of cloud computing in management and education, *Materials Today:* Renewable Energy Assessment Using Novel Ensembling Approach. *Journal of Nano-and Electronic Physics*, *15*(4).

[16] William, P., Shrivastava, A., Shunmuga Karpagam, N., Mohanaprakash, T.A., Tongkachok, K., Kumar, K. (2023). Crime Analysis Using Computer Vision Approach with Machine Learning. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_25

[17] R. Detrano, A. Janosi,W. Steinbrunn, M. Psterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, ``International application of a new probability algorithm for the diagnosis of coronary artery disease,'' Amer. J. Cardiol., vol. 64, no. 5, pp. 304310, Aug. 1989.

[18] J. H. Gennari, P. Langley, and D. Fisher, ``Models of incremental concept formation,'' Artif. Intell., vol. 40, nos. 13, pp. 1161, Sep. 1989.

[19] Y. Li, T. Li, and H. Liu, ``Recent advances in feature selection and its applications,'' Knowl. Inf. Syst., vol. 53, no. 3, pp. 551577, Dec. 2017

[20] Baashar Y, Gamal A, Alhussian H, et al. (2022) Effectiveness of artificial intelligence models for cardiovascular disease prediction: network meta-analysis. *Computational Intelligence and Neuroscience* 2022: 1–12. Article ID 5849995, 12.

[21] Chowdhury MNR, Ahmed E, Siddik MAD, et al. (2021) *Heart Disease Prognosis Using Machine Learning Classification Techniques. 2021 6th International Conference for Convergence in Technology*, pp. 1–6.

[22] Ghosh P, Azam S, Jonkman M, et al. (2021) Efficient Prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access* 9: 19304–19326.

[23] Islam S, Jahan N, Khatun ME (2020) *Cardiovascular Disease Forecast Using Machine Learning Paradigms. 2020 Fourth International Conference on Computing Methodologies and Communication*. ICCMC), pp. 487–490.

[24] Kavitha M, Gnaneswar G, Dinesh R, et al. (2021) *Heart Disease Prediction Using Hybrid Machine Learning Model. 2021 6th International Conference on Inventive Computation Technologies*. ICICT), pp. 1329–1333.

[25] Khan SI, Choubey SB, Choubey A, et al. (2022) Automated glaucoma detection from fundus images using wavelet-based denoising and machine learning. *Concurrent Engineering* 30(1): 103–115.

[26] Nagavelli U, Samanta D, Chakraborty P (2022) Machine Learning Technology-Based Heart Disease Detection Models. *Journal of Healthcare Engineering* 2022: 1–6.