Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Empowering Health Prognosis: Quine Mccluskey Binary Classifier Algorithm for Predicting Heart Disease

Shanthi Makka[1], S. Bharath Reddy[2], Preeti Jaidka[3], R. Raja Kumar[4], Asma Begum[5]

[1]*Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, Telangana, India, dr.shanthimakka@gmail.com*
[2]*Department of CSE, KG Reddy College of Engineering and Technology, Chilukuru village Moinabad R R Dist Telangana, India, s.bharathreddy@kgr.ac.in*
[3]*JSS Academy of Technical Education Noida, U.P, India, preetijaidka@jssaten.ac.in*
[4]*CSE Department, RGMCET, Nanadyal, Andhra Pradesh, India, rajakumar.rajaboina@gmail.com*
[5]*Vardhaman College of Engineering, Hyderabad, Telangana, India, asma1512@vardhaman.org*

Numerous researchers employed the Quine McCluskey Binary Classifier (QMBC) for rapid identification of cardiac patients. These integrated models, including decision trees, random forests, K-nearest neighbors, naive Bayes, support vector machines, and multilayer perceptron's, demonstrate effective performance on binary classification datasets. Use chi-square, analysis of variance, and principal component analysis to find and select features for predictive modeling. The ensemble model uses the Minimum Boolean expression from the seven models, works better than the best methods that are currently available. The collection includes examples from Cleveland HD, HD Comprehensive, and CVD. Its main goal is to improve accuracy by using advanced ensemble methods. In addition, a more in-depth study looks into how to combine Voting Classifier and Stacking Classifier, with the goal of reaching 100% accuracy. This paper adds to the field by introducing a strong method called QMBC and showing how strategic ensemble methods can be used to improve performance even more.

Keywords: ANOVA, Quine McCluskey Binary Classifier; chi-square; Ensemble approach; Machine learning; principal component analysis.

1. Introduction

HD is a broad term for a variety of heart and blood vessel diseases. Arterial diseases and heartbeat issues fall into this category. They use the patient's medical history (Makka, S., et al., 2021) and blood sugar, cholesterol, and blood pressure tests to determine it. Modern (Makka, S., et al., 2022, Sunitha, L.et al., 2022) tests for heart problems include X-rays, ultrasounds, arterial angiography, radioactive tests, MRI scans, and CT scans. Heart failure is caused by long-term problems that hurt or weaken the heart muscles, which lowers the ejection fraction. The disease can hurt other important body parts very badly and can happen to both adults and children. Age, race, family background, genetics, living choices, and having a cardiovascular disease (CVD) or genetics that was there before are the main Factors that increase your risk of heart problems failure. ML is becoming a more and more important tool for finding diseases early on. Its goal is to find patterns in events and come to conclusions that make sense in light of new knowledge. ML is becoming a more and more important tool for finding diseases early on. Its goal is to find patterns in events and come to conclusions that make sense in light of new knowledge. Researchers have looked into how to combine different techniques to make mixed models that work better than solo models. Most of the time, these types have two stages Feature Extraction (FE) and Feature Selection (FS) are used to choose a group of traits in phase-1. Then, the group is fed into the algorithms that were used in step 2. Heart disease records often have a lot of different traits, some of which are important, some of which are not, and sometimes there are even similar characteristics. Relevant attributes change how the target class is defined, while useless attributes don't change how the output class is described. On the other hand, redundant characteristics add noise to the description of the target class instead of adding any new information (S. Shilaskar and A. Ghatol, 2013). To make the classification models better, it is important to get rid of traits that Not just altering the classification outcomes, but also causing a slowdown in the system. There are features in the datasets that aren't useful or are repeated, which adds to the noise and doesn't tell us anything about the target class. This is why the HD diagnosing system needs dimensionality reduction or FS methods. You are less likely to over fit, the model can generalize better, predictions are more accurate, and you need to do less work, so there are fewer features (N. C. Long, et al., 2015). So that the forecasting process goes faster, we use algorithms to find and pick out features. We use the ANOVA and Chi-Square tests to pick out the top 10 traits and divide the dataset into smaller parts. After that, we use Principal Component Analysis to find nine prime components in the set. To find the smallest true or false statement for the goal feature, we use the Quine McCluskey method on a group of all seven models. A novel hybrid QMBC technique effectively discerns between patients with and without heart disease. It incorporates Naive Bayes; SVM, LR, DT, RF, MLP, and K NN methods are among the QMBC models. It is fairly proficient at managing collections of binary categorical data.

2. Literature Survey

This paper aimed to develop a model utilizing machine learning to forecast mortality in individuals diagnosed with coronary artery disease by incorporating additional health factors. Data from over 13,000 CAD patients who attended cardiac rehabilitation between 1995 and 2016 was analyzed, revealing age and peak metabolic equivalents as key predictors of mortality (C. G. D. S. E. Silva, et al., 2022). The model, using the survival tree algorithm, demonstrated a high level of accuracy in predicting death over a long-term follow-up period. The ST algorithm divided patients into eight groups with different chances of survival, showing that maximal exercise capacity is a key factor in prognosis. This machine learning-based method worked better than traditional clinical variables, showing that age and exercise capacity are strong indicators of CAD death. The study comes to the conclusion that machine learning techniques, specifically the ST algorithm, offer a new and useful way to predict death, providing doctors with additional insights to enhance their decision-making process for CAD patients. It utilizes the CART algorithm, a supervised machine learning technique, to create a model and make predictions related to heart disease, a significant issue worldwide. Heart diseases cause about one-third of all deaths in the world, so the study is mostly about how to improve early detection in medicine by using advanced machine learning applications. The CART algorithm is used to figure out how to make decisions and predict heart disease by revealing complex connections between input and output variables. The study lists the factors that have the most significant effects on heart disease and ranks their level of importance. The model makes predictions that are an impressive 87% of the time, which proves that it is reliable across a wide range of performance parameters. It is important to note that the extracted decision rules provide a simplified framework for clinical applications, which helps doctors and patients who are limited by time and cost when diagnosing and treating heart disease. In conclusion, this study presents a strong CART-based algorithm that not only helps doctors but also solves problems that patients actually face while getting diagnosed with and treated for heart disease. This paper explores how machine-learning algorithms (M. M. Nishat, et al., 2022) can be used to improve patient survival rates for heart failure by addressing imbalances in datasets. By utilizing a variety of supervised algorithms and techniques such as SMOTE-ENN and hyper parameter optimization, the study found that the Random Forest Classifier demonstrated superior performance with 90% test accuracy. This study introduces machine-learning algorithms (P. Ghosh, S. et al., 2021). Relief and LASSO feature selection methods for cardiovascular disease (CVD) prediction. Early detection of CVD is important to prevent and reduce mortality. Improving data collection and prioritization processes can increase the accuracy of training models. Data and statistical data from various locations including Cleveland, Long Beach, Virginia, Switzerland, and Hungary were used. Recovery and LASSO feature selection methods improve predictions. Hybrid classifiers perform better in model training. In addition to traditional classifiers, packaging and transportation systems (S. Bashir, et al., 2019) also used.

Table 1. Literature Survey

Title of the Paper	Authors, Year of Publication	Outcomes
Prediction of mortality in coronary artery disease: Role of machine learning and maximal exercise capacity	C. G. D. S. E. Silva et al., 2022	Additional subdivisions within each patient subgroup are determined by age or peak METs cut points.
A classification and regression tree algorithm for heart disease modeling and prediction	M. Ozcan and S. Peker, 2023	The algorithm aids both healthcare professionals and patients facing Constraints related to cost and time in diagnosing and treating heart disease.
A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyper parameter optimization for imbalanced heart failure dataset	M. M. Nishat, et al., 2022	Enhancing machine learning algorithm performance is accomplished by employing the SMOTE-ENN algorithm and hyper parameter optimization techniques.

Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques	P. Ghosh, S. et al., 2021	The greatest accuracy was achieved when employing RFBM and Relief feature selection techniques.
Heart disease identification method using machine learning classification in healthcare	J. P. et al., 2020	The suggested diagnostic system (FCMIM-SVM) demonstrated notable accuracy when compared to previously proposed methods.

A range of metrics, including accuracy, precision, error rate, F1 score, and false positive rate, were analyzed. The model achieved a peak accuracy of 99.05% using RFBM and Relief feature selection methods. It can predict heart disease well. This paper uses a combination (A. K. Gárate-Escamila, et al., 2020) of PCA (principal analysis) and feature selection (McSweeney, et al., 2016) to improve the prediction of cardiovascular disease. The researchers tested six different machine learning methods using data sourced from the UCI Machine Learning Repository, which includes 74 features and labels indicating the presence or absence of heart disease. They introduced a new size reduction method called CHI-PCA, which looks specifically at factors associated with cardiovascular disease. The results show that CHI-PCA outperforms other methods, especially in random forest. The prediction accuracy for cardiovascular disease was an accuracy of 98.7% was achieved for the Cleveland dataset, 99.0% for the Hungary dataset, and 99.4% for the combined Cleveland-Hungary dataset. Features considered include cholesterol level, maximum heart rate, chest pain, and ST depression and symptoms, and cardiovascular system metrics were deemed anatomically and physiologically significant and were analyzed using chi-square feature selection (J. P. Li et al., 2020). The combination of chi-square and PCA outperforms other classification methods in tests. Interestingly, performing PCA directly on raw data results in lower accuracy; this indicates that CHI-PCA is a better method for predicting heart disease and also reduces risk.

3. Machine Learning Models

3.1 System Architecture

Systematic designs using machine-learning models (M. Ayar, A. et al., 2022) to accurately and effectively predict cardiovascular diseases. Use classification techniques such as Chi-Square, Analysis of Variance, and PCA, as well as SVM, Naive Bayes, Decision Tree, Random Forests, K-Nearest Neighbor, and Multilayer Perception to extract dataset behaviour to refine the model. Combining these methods does not affect the product, but reduces the size, speeds up the calculation and makes it better. New Quine McCluskey Binary Classifier (QMBC) integration technology provides predictive results. The system also has a Voting Classifier that includes all base models and additions like a Voting Classifier with Adaboost + RF and a Stacking Classifier with RF + MLP using LightGBM.

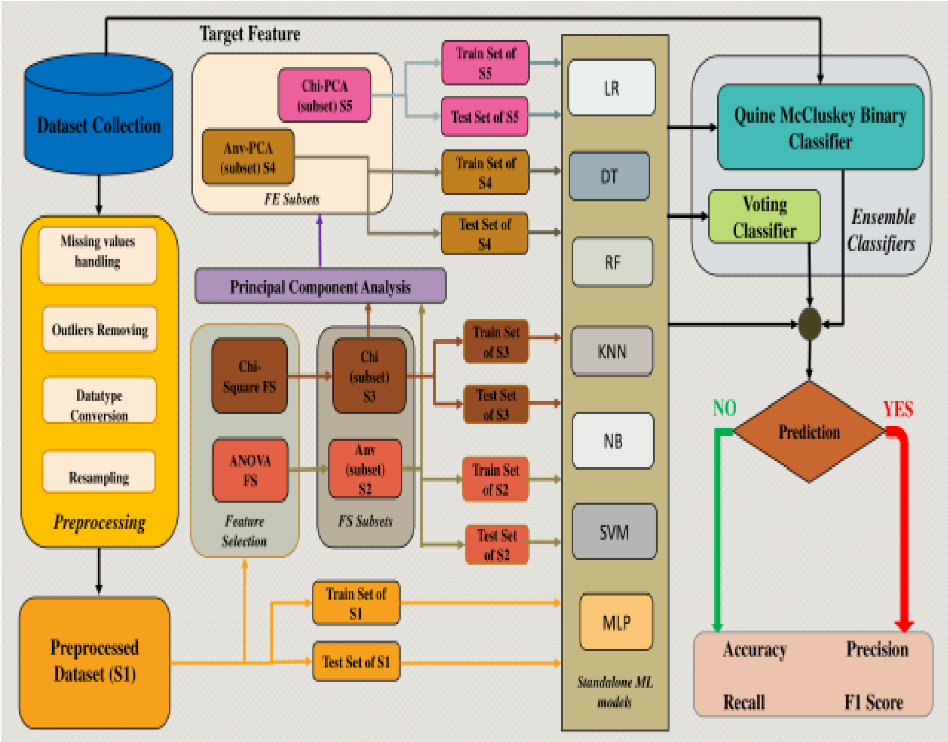

Figure 1. System Architecture

The architecture is flexible and has been tested on different datasets, including Cleveland HD, comprehensive HD, and CVD. It may be possible to get higher accuracy by using ensemble methods, which makes it a strong and flexible way to predict heart disease (I. D. Mienye, et al., 2020). Overall, the research highlights the importance of using advanced algorithms and methods in predicting heart failure survival.

3.2 Algorithms

Random Forest: Leo Breiman and Adele Cutler invent this algorithm and it combines output of decision tree. It is popular because it is easy to use and can classify and regress.

Logistic Regression: This machine learning classification algorithm estimates class likelihood based on dependent variables. In summary, the logistic regression model adds up the features (usually a bias term) and calculates the logistic result.

Decision tree: The structure is similar to a flowchart, with each node representing a test of a specific attribute. The branches show the different results of the test, and the leaf nodes represent the decision made based on all the attributes.

K-Nearest Neighbors (KNN) Algorithm: This supervised learning classifier, categorizes data points based on their proximity to one another without making any assumptions about their underlying distribution.

Naive Bayes: Based on Bayes' Theorem and assuming predictors are independent groups or things. Naive Bayes classifiers believe that presence of a single feature within a class does not impact its presence of others.

SVC: SVC is a supervised machine-learning algorithm used to group things. SVC locates the optimal hyper plane to divide data points into two groups by mapping them to a multidimensional space.

MLP: A modern artificial neural network, known as a feed forward network, consists of interconnected neurons with a nonlinear activation function. It is organized into multiple layers and can effectively separate data that cannot be separated by a straight line. The term "MLP" is not accurate as the first perceptron used a different type of activation function called the Heaviside step function, unlike the nonlinear activation function used in modern networks.

Stacking Classifier: This ensemble method sends the results of several classifiers to a meta-classifier for classification.

Voting Classifier: A voting classifier is a type of machine learning model that generates a final prediction by aggregating the predictions of several base models. The classifier's final output may be enhanced through modification of how the predictions are merged.

4. Experimental Approach

4.1 Dataset collection

This study predicts coronary heart disease using three free benchmark datasets. Many research studies use the University of California (UCI) Cleveland dataset. There are 76 attributes, but researchers usually focus on 14. The second dataset, HD (Comprehensive), includes patient records from various locations and contains important variables related to heart health. The third dataset, CVD, is the largest and consists of a large number of patient records with relevant attributes. By analyzing (C. B. C. Latha and S. C. Jeeva, 2019) these datasets, researchers, and doctors can identify Factors posing risks and indicators for diagnosis and treatments for heart disease, which can significantly improve prevention and treatment efforts.

4.2 Data processing

The initial phase of the data processing pipeline involves the use of a panda’s data frame to organize and work with the dataset effectively. Then, columns that aren't needed are removed to make the dataset smaller, which speeds up computations and lets the computer focus on the important features. Through recognizing of the most significant patterns, PCA is used to reduce the number of dimensions in data. After that, seaborne and matplotlib are used to show the data, which gives information about how the data is distributed, how it is related to other data, and possible patterns. Label Encoding is done with Label Encoder, which turns categorical variables into numerical format so that machine-learning models can work together. Choosing the right features is an important step. The process includes Anova FS (which uses analysis of variance to figure out how important a feature is) and Chi2 FS (which uses the chi-square statistic). For comparison, the original dataset is kept. This all-around approach to processing data guarantees a clean and useful dataset that is ready to be analyzed and modeled later on. Dimensionality reduction, visualization (H. Li, et al., 2018), label encoding, and feature selection all work together to make the data better and more suitable for predictive modeling tasks.

4.3 Training & Testing

Machine learning requires data in utilizing training and testing datasets for evaluating model effectiveness. After data preparation, it is split into feature variables (X) and target variables (y). The input data is contained in the feature variables of X, and the labels or outcomes are contained in y. About 70% to 80% of the dataset is usually set aside for the training set (X_train, y_train). This gives the machine-learning model a lot of examples to learn from and find patterns in. During the training phase, the model changes its settings based on the features and labels that go with them. This makes it better at making accurate predictions. The last 20–30% is the testing set (X_test, y_test), which is a separate set of data that the model hasn't seen before during training. This separation makes sure that the model's generalization performance can be judged fairly, which measures how well It possesses the ability to generate precise predictions for new data which does not seen before. Checking the model on this specific dataset helps find problems with over fitting and shows how it can be used in the real world.

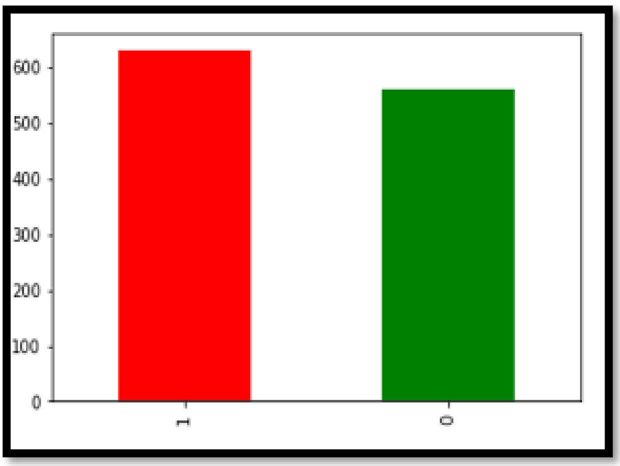

Figure 2. Cleveland dataset accuracy

Finding the right balance between training and testing data (Makka, S., et al., 2022) is important for creating a strong, well-generalized machine-learning model that can make accurate predictions about a wide range of new situations. This division makes sure that the model works well and is reliable when used in real life.

Cleveland Dataset: The experimental environment is created using a dataset sourced from the Kaggle machine-learning repository. In the dataset, we found 297 instances.

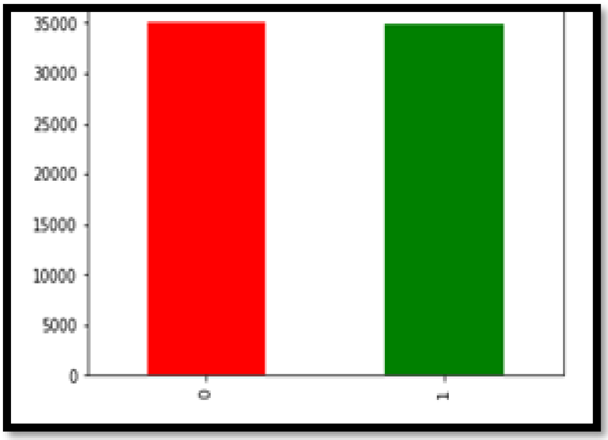

Figure 3. CVD dataset accuracy

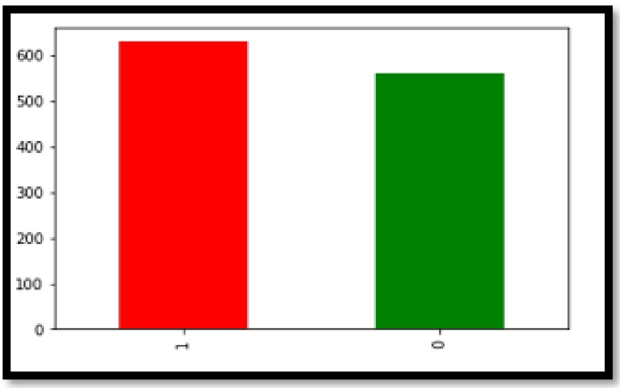

Figure 4. HD comprehensive dataset accuracy

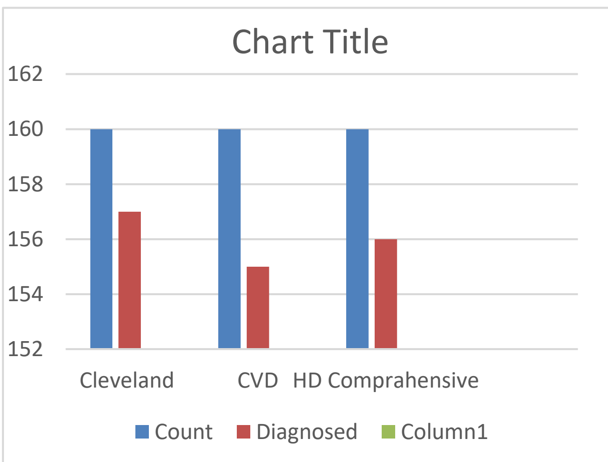

Figure 5. Overall three datasets Accuracy

5. Proposed Approach and Methodology

A novel Quine McCluskey binary classifier (QMBC) method has been developed to differentiate heart patients during the process. It employs regression, decision trees, random forests, K-nearest neighbors, naive Bayes, support vector machines, and multilayer perceptron models. Combining Chi-Square, ANOVA, and PCA improves the model (R. K. Sevakula and N. K. Verma, 2017). This completes a plan that reduces dimensions, speeds computation, and optimizes performance. QMBC ensemble technique combines model results to make an accurate prediction.

A number of different classifiers are built into the model architecture, including a Voting Classifier that uses all base models and an extended ensemble approach (Voting Classifier with Adaboost + RF and Stacking Classifier with RF + MLP using LightGBM). Three datasets are used for evaluation: Cleveland HD, comprehensive HD, and CVD.

The existing survey says that QMBC is 98% accurate. As an addition, the system looks into other ensemble methods, such as Voting and Stacking Classifiers, in the hopes of improving performance even more and possibly reaching 100% accuracy, which would help heart disease prediction models, move forward.

5.1 Significance of QMBC

The Quine McCluskey Binary Classifier is a method used for simplifying Boolean functions, primarily in digital circuit design. It is named after Willard Van Orman Quine and Edward J. McCluskey, who created it in the early 1950s. The significance of QMBC lies in its ability to simplify Boolean functions, leading to more efficient digital circuit designs with reduced complexity, cost, and power consumption.

Boolean Function Simplification: QMBC is a powerful tool for simplifying Boolean functions, diminishing the quantity of terms or literals required to act a function. This simplification can lead to more efficient digital circuits with fewer gates, resulting in cost savings, faster operation, and reduced power consumption.

Applications in Digital Design: QMBC finds extensive applications in digital circuit design, where Boolean functions are fundamental for describing the behavior of logic gates and circuits. By simplifying these functions, engineers can optimize the design of digital systems, making them more reliable and easier to manufacture.

Minimization of Logic Circuits: One of the primary goals of QMBC is to minimize the number of logic gates needed to implement a given Boolean function. This reduction process aids in decreasing the complexity and size of digital circuits, making them more manageable and efficient.

Complexity Reduction: Boolean functions can become very complex, especially in large-scale digital systems. QMBC provides a systematic approach to reducing this complexity by identifying redundant terms and combining them to form simpler expressions.

Automation: While QMBC can be done manually, it's often automated through software tools. These tools take the Boolean function as input and apply the Quine-McCluskey algorithm to generate the simplified expression automatically. This automation saves time and reduces the chances of human error.

Compatibility: QMBC is compatible with various other techniques and algorithms used in digital design, such as Karnaugh maps and Espresso heuristic logic minimization. Engineers often use a combination of these methods to achieve the best possible simplification for a given Boolean function.

6. Results and Discussion

Accuracy: The measure of a test's ability to distinguish between individuals who are sick and those who are healthy is termed its accuracy. To gauge the accuracy of a test, it's essential to determine the percentage of cases that are true positives and true negatives.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: Precision refers to the proportion of accurately classified instances or samples in relation to those that were correctly identified as positives. Therefore, here is the formula to calculate precision.

$$\text{Precision}=\frac{True\ Passitive}{True\ Passitive+False\ Passitive} \quad (2)$$

Recall: In machine learning, recall serves as a metric indicating the model's capability to identify all pertinent instances of a specific class, showcasing its proficiency in capturing such instances. It's computed by dividing the ratio of correctly predicted positive observations to the total number of actual positives.

$$\text{Recall}=\frac{TP}{TP+FN} \quad (3)$$

F1-Score: The F1 score is a method for evaluating the accuracy of a machine-learning model, which combines its precision and recall scores. The accuracy metric indicates how often a model made correct predictions throughout the entire dataset.

$$\text{F1 Score}=\frac{2}{\left(\frac{1}{Precision}+\frac{1}{Recall}\right)} \quad (4)$$

$$\text{F1 Score}=\frac{2\times Precision\times Recall}{Precision+Recall} \quad (5)$$

Analysis and comparative study of above-mentioned features on various algorithms is depicted through following graphs.

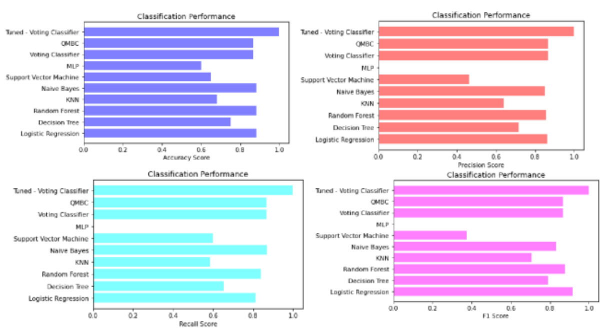

Figure 6. Performance Comparison Graphs Cleveland HD Dataset

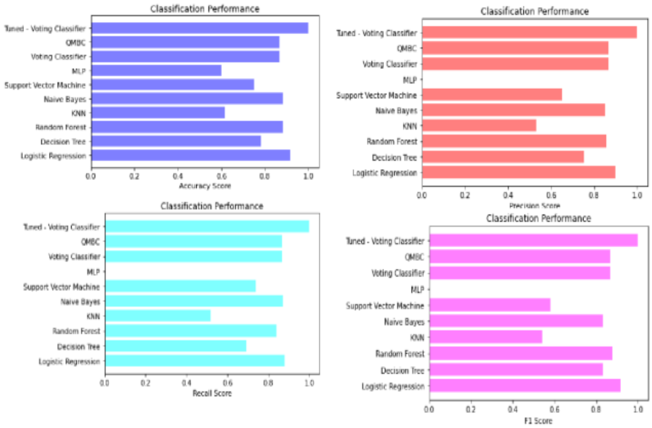

Figure 7. Performance Comparison Graphs Cleveland HD Dataset for Anova –FS

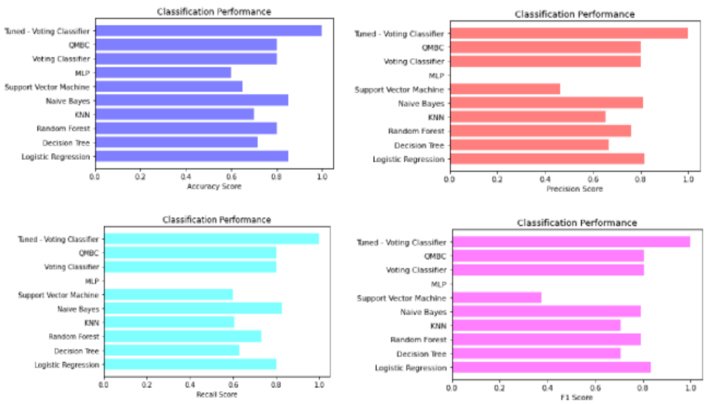

Figure 8. Performance Comparison Graphs Cleveland HD Dataset for Chi2 FS

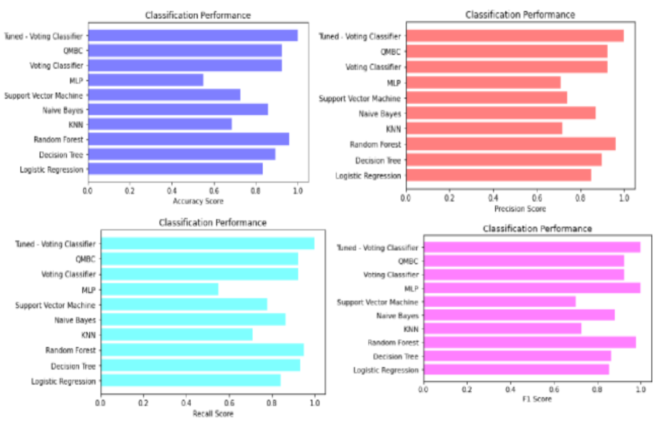

Figure 9. Performance Comparison Graphs Comprehensive HD Dataset

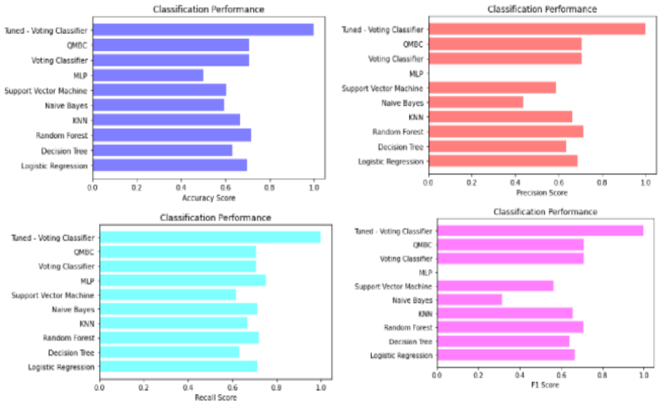

Figure 10. Performance Comparison Graph for CVD Dataset

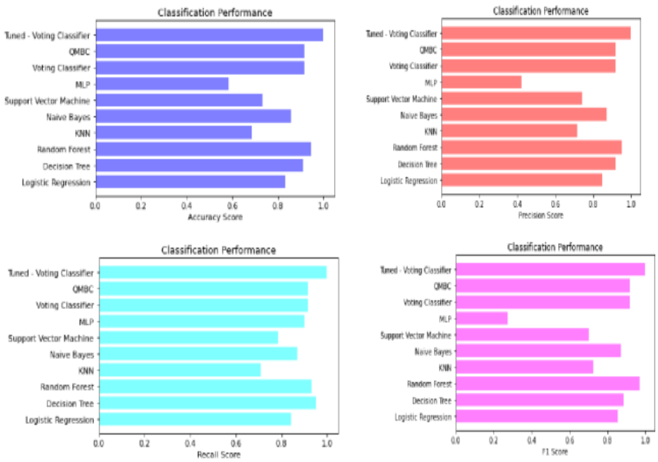

Figure 11. Performance Comparison Graphs Comprehensive HD Dataset for Anova –FS

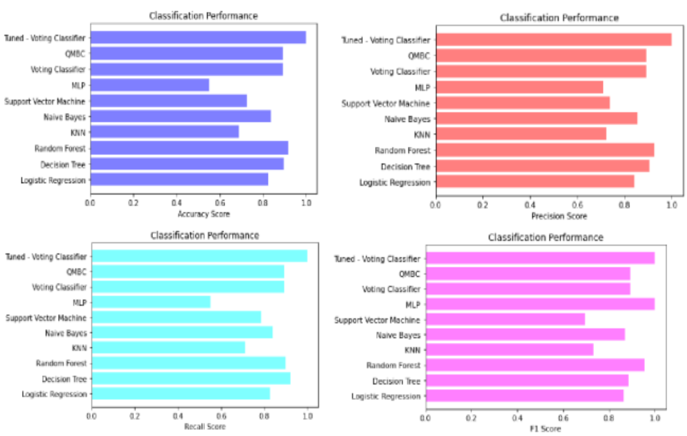

Figure 12. Performance Comparison Graphs Comprehensive HD Dataset for Chi2 FS

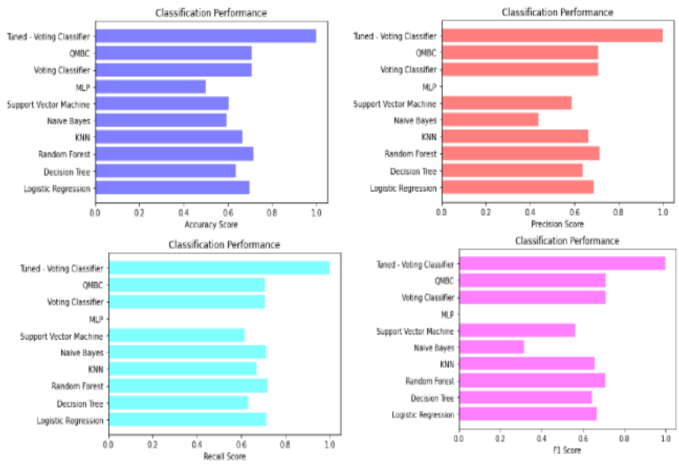

Figure 13. Performance Comparison Graphs CVD Dataset for Anova –FS

7. Conclusion

This paper emphasizes on how well standalone machine learning models worked and created a new QMBC to predict heart disease across a variety of datasets. The datasets were carefully prepared before being used in models. This was done using Chi-Square, Anova, and PCA techniques. The QMBC model, which used a combination of ANOVA and PCA feature extraction, was more accurate, precise, recallable, and had a higher f1-score than other methods. The amazing outcomes in the Cleveland, cardiovascular, and HD (Comprehensive) datasets show that the suggested method works. Enhancing the precision of predictions for heart disease in the future, researchers will work on datasets that aren't balanced and explore deep learning techniques. Extending ensemble techniques like Voting Classifier and Stacking Classifier could lead to higher accuracy, which could lead to better diagnostic tools for heart health. This study improves heart disease prediction and opens the door to machine learning model research and improvement to save lives.

References

1. A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, ''Classification models for heart disease prediction using feature selection and PCA,'' Informat. Med. Unlocked, vol. 19, 2020.
2. C. B. C. Latha and S. C. Jeeva, ''Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,'' Informat. Med. Unlocked, vol. 16, Jan. 2019, Art. no. 100203.
3. C. G. D. S. E. Silva, G. C. Buginga, E. A. D. S. E. Silva, R. Arena, C. R. Rouleau, S. Aggarwal, S. B. Wilton, L. Austford, T. Hauer, and J. Myers, ''Prediction of mortality in coronary artery disease: Role of machine learning and maximal exercise capacity,'' Mayo Clinic Proc., vol. 97, no. 8, pp. 1472–1482, Aug. 2022.
4. H. Li, Y. Cui, Y. Liu, W. Li, Y. Shi, C. Fang, H. Li, T. Gao, L. Hu, and Y. Lu, ''Ensemble learning for overall power conversion efficiency of the allorganic dye-sensitized solar cells,'' IEEE Access, vol. 6, pp. 34118–34126, 2018.

5. I. D. Mienye, Y. Sun, and Z. Wang, ''Improved sparse autoencoder based artificial neural network approach for prediction of heart disease,'' Informat. Med. Unlocked, vol. 18, Jan. 2020, Art. no. 100307.
6. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, ''Heart disease identification method using machine learning classification in ehealthcare,'' IEEE Access, vol. 8, pp. 107562–107582, 2020.
7. M. Ayar, A. Isazadeh, F. S. Gharehchopogh, and M. Seyedi, ''Chaoticbased divide-and-conquer feature selection method and its application in cardiac arrhythmia classification,'' J. Supercomput., vol. 78, pp. 5856–5882, Mar. 2022.
8. M. M. Nishat, F. Faisal, I. J. Ratul, A. Al-Monsur, A. M. Ar-Rafi, S. M. Nasrullah, M. T. Reza, and M. R. H. Khan, ''A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset,'' Sci. Program., vol. 2022, pp. 1–17, Mar. 2022.
9. M. Ozcan and S. Peker, ''A classification and regression tree algorithm for heart disease modeling and prediction,'' Healthcare Anal., vol. 3, Nov. 2023, Art. no. 100130.
10. Makka, S., Arora, G., & Mopuru, B. (2021, November). IoT based health monitoring and record management using distributed ledger. In Journal of Physics: Conference Series (Vol. 2089, No. 1, p. 012030). IOP Publishing.
11. Makka, S., Arora, G., Reddy, S. S. T., & Lingam, S. (2022, November). Use of Machine Learning Models for Analyzing the Accuracy of Predicting the Cancerous Diseases. In International Conference on Innovations in Data Analytics (pp. 169-180). Singapore: Springer Nature Singapore.
12. Makka, S., Sreenivasulu, K., Rawat, B. S., Saxena, K., Rajasulochana, S., & Shukla, S. K. (2022, December). Application of Blockchain and Internet of Things (IoT) for Ensuring Privacy and Security of Health Records and Medical Services. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 84-88). IEEE.
13. McSweeney, J. C., Rosenfeld, A. G., Abel, W. M., Braun, L. T., Burke, L. E., Daugherty, S. L., ... & Reckelhoff, J. F. (2016). Preventing and experiencing ischemic heart disease as a woman: state of the science: a scientific statement from the American Heart Association. Circulation, 133(13), 1302-1331.
14. N. C. Long, P. Meesad, and H. Unger, ''A highly accurate firefly based algorithm for heart disease prediction,'' Expert Syst. Appl., vol. 42, no. 21, pp. 8221–8231, Nov. 2015.
15. P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, ''Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques,'' IEEE Access, vol. 9, pp. 19304–19326, 2021.
16. R. K. Sevakula and N. K. Verma, ''Assessing generalization ability of majority vote point classifiers,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 12, pp. 2985–2997, Dec. 2017.
17. S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, ''Improving heart disease prediction using feature selection approaches,'' in Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2019, pp. 619–623.
18. S. Shilaskar and A. Ghatol, ''Feature selection for medical diagnosis: Evaluation for cardiovascular diseases,'' Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.
19. Sunitha, L., Makka, S., Madhu, S., & Bheemeswara Sastry, J. (2022). Study on influence of outliers on the performance of various classification algorithms. In Innovations in Electronics and Communication Engineering: Proceedings of the 9th ICIECE 2021 (pp. 437-445). Singapore: Springer Singapore.