Ensemble Methods Based Feature Selection For Cardiovascular Diseases

Md Sirajul Huque[1*], V. Kiran Kumar[2]

[1*,2]Dept of Computer Science and Technology, Dravidian University, Andhra Pradesh, India,
[1]sirajulhuque1985@gmail.com, [2]kirankumar.v@rediffmail.com

***Corresponding Author:** Md Sirajul Huque
*Dept of Computer Science and Technology, Dravidian University, Andhra Pradesh, India,
sirajulhuque1985@gmail.com

Cite this paper as: Md Sirajul Huque, V. Kiran Kumar (2024) Ensemble Methods Based Feature Selection For Cardiovascular Diseases*, Frontiers in Health Informatics*, 13 (4), 262-272

Abstract: In the healthcare sector, developing the feature selection model to predict cardiovascular diseases (CVD) is a challenging task. The potential risk factors and predictors and carefully considering feature selection techniques are necessary to create dependable prediction models. This paper explores the significance of feature selection approaches in enhancing the CVD prediction models. Although traditional machine learning models with individual methods have been used for feature selection, but their performance is not so much appropriate for the various data sets, specifically healthcare datasets. Thus, we have taken ensemble machine learning methods to solve such issues for the healthcare dataset. We have considered the ensemble machine learning model based on the combination of two methods, such as the combination of support vector machine (SVM) and random forest (RF) as (SVM+RF), decision tree (DT) and random forest (RF) as (DT+RF). We demonstrated our model using the CVD dataset, and it performed better than traditional methods. Our model is more effective in selecting the correct features from the data set and predicting the CVD.

Keywords: Cardio Vascular Diseases (CVD), Feature Selection, Lasso Regularization, Backward Elimination, Machine Learning Models.

1 Introduction

Cardiovascular diseases refer to a group of conditions that affect the heart and blood vessels. These can harm the human cardiovascular system in several ways. When a patient has cardiac problems, they fall into two categories. Globally, cardiac arrest is among the leading causes of death. It takes a life every few seconds. These heart conditions affect at least half of the US population. Many variables are taken into account while estimating the risk of cardiac arrest fatalities. Hypertension, dyslipidemia, alcoholism, tobacco use, poor eating habits, and inactivity (including yoga and meditation) are a few of these. A typical collection of symptoms might indicate any of the following conditions: Chest pain and other heart-related symptoms [1][2]. Cardiovascular disease risk factors included inactivity, smoking, heavy alcohol use, high blood pressure, heredity, gender, age, cholesterol, and poor dietary habits. A poor diet, inactivity, alcohol use, and cigarette smoking are all factors that raise the risk of cardiovascular disease. Analysts employ a tiered system of numerical data mining techniques to evaluate cardiac disease [3].

Cardiovascular disease (CVD) diagnosis, treatment, and prognosis depend heavily on machine learning (ML). A person's age, lifestyle, medical history, and other risk factors are among the many patient data points that machine learning algorithms scour to identify patterns and predictors of cardiovascular disease (CVD) [4]. These models search medical images for signs of cardiac issues or arterial plaque, facilitating early diagnosis and prompt treatment. Machine learning (ML) assesses patient data using genetic predispositions and medical history to provide individual treatment recommendations that maximize medication efficacy. Additionally, technology can now more accurately identify vital signs in real time thanks to machine learning (ML)[5][1], allowing medical professionals to react quickly to any abnormalities. To fully exploit ML integration's potential and eliminate biases and risks, it's critical to use it carefully and purposefully [6].

Although different methods have been developed for CVD diagnosis using ML algorithms, hybrid or ensemble models are rarely used to test CVD. Thus, the ensemble model of two methods is considered to evaluate the CVD dataset and compare its evaluation performance with individual methods. This work's main objective is to find relevant methods to make ensemble models and perform well compared to others. Searching for relevant methods to form ensemble models is challenging for evaluating carefully using CVD dataset.
The following are a few of the crucial contributions made to the paper:

- To emphasize essential feature selection techniques to create trustworthy CVD prediction models.
- To gain further knowledge about feature selection techniques, our ensemble model evaluates their effectiveness, interpretability, and computing efficiency.
- To develop prediction models for CVD by employing several machine learning ensemble approaches such as SVM+RF, DT+RF, LR+RF, etc.
- To train these models using datasets produced by feature extraction using our ensemble model and the raw data set.

The remaining sections of the paper are outlined as follows. In section 2, we have elaborated on the background of this paper. Section 3 is considered for the proposed ML model with data processing. Section 4 explains the experimental evaluation with different performances. The whole paper is concluded in section 5.

2 Related Work

Most of the research work has been used "UCI Cleveland Heart Disease Data Set." There are fourteen unique characteristics in the data collection. The dataset is divided into two parts: training data and testing data. Their preferred feature option was for data pre-processing. They employed a traditional scaler for a scattered range to categorise the data. Several classification models are used, such as K-NN, SVM, and RF. They have discovered the accuracy of each model. K-NN and SVM are the most accurate models for predicting the existence of cardiovascular disease, they found after extensive investigation [7]. To enhance the performance of the models, they employed hyperparameters such as cross-validation and grid search. It has been demonstrated that using logistic regression to predict CVD is a suitable approach [8]. The result of the data pre-processing was that they had carried out data cleaning, which included the removal of duplicate values. Finally, they found that Random Forest is a good predictor of cardiac issues [9]. Several sources provided the data for this study. After pre-processing, they found that just 6 of the 303 patient records included the required information. In this dataset, 297 records were part of the pre-processing findings, 137 had cardiac abnormalities, and 160 did not.
We used several machine learning techniques, such as NB, LR, DT, KNN, etc. The dataset was split into training and testing subsets, which made up 30% and 70% of the total. They ascertained the degrees of Sensitivity, Precision, and Accuracy. They discovered that the most effective model for predicting the risk of heart disease was a hybrid one that included SVM and Naïve Bayes. [10] Two sources provided the data set utilised in this article: IEEE data port and Kaggle. The final result includes 2213 submissions with 12 characteristics in total. Information Pre-processing was done to get rid of the missing values. Training and testing were done using data in a 70% to 30% ratio. K-Fold cross-validation was applied to the training data sets. For both training and analysis, models including RF, KNN, SVM, LR, CART, Ada Boost, and Naïve Bayes were employed. Finally, they found that Random Forest is a good predictor of cardiac issues. [11]. To gather information for this research, 14 distinct characteristics were used. KNN and Random Forest were their preferred models. Making predictions is the primary goal of this study. The accuracy level, often stated as a percentage, is the outcome. In the end, it was demonstrated that the K-NN model best-predicted outcomes.[12].
The 1025 patient records that comprise this article's data set were acquired via Kaggle. Afterwards, feature selection during pre-processing was done using the suggested PCHF approach. The dataset was then divided in half, with 80% designated for training and 20% for testing. Nine ML models were employed. They discovered 526 people with cardiac disease and 499 people in excellent condition. They finally decided that the Decision Tree was the best method for predicting heart failure illness after considerable consideration. The accuracy data was also calculated for precision, recall, F1-score, and K-fold cross-validation. The Cleveland Heart Disease Data Set, which has 1025 items and 14 characteristics, was used [13]. Eighty percent of the data were used in the training set, while twenty percent were used in the testing set. They used Naïve Bayes, Random Forest, K-NN, and Support Vector Machine as their four machine learning models. Random Forest produced the most outstanding results when we compared the algorithms [14][15]. They examined the UCI data set used in this paper. Machine learning algorithms were heavily utilised.

The F1 score, accuracy, precision, and recall were established. After careful consideration, they concluded that Hybrid Classification performed more accurately than K-NN [16,17,18]. Two different types of information were evaluated in this study. The "Cardiovascular Disease Data Set" has twelve features, while the "Heart Disease Data Set" has thirteen. Out of the data set, 80% was utilised for training, and only 20% was used for testing. K-fold cross-validation was used to calculate the Silhouette Score, Accuracy, Precision, and Recall. Ultimately, they discover that Random Forest performs better than the other machine learning algorithms [19]. Bhuyan and et al., suggested the single-setting disease analysis in [21,22].
From the above research approaches, we observed that the hybrid model or ensemble model is not so much developed by the different researchers to improve the experimental performance, which was a challenging task to encourage us and proceed to a new ensemble model to compare with an existing individual model which is explained in subsequent sections of this paper.

3 Proposed Methodology

While machine learning models employ several techniques for feature selection, each method is tailored to a particular dataset. We have taken into account the ensemble model for feature selection. To obtain the final findings from the dataset, as seen in Figure 1, we have thought about prepping the data using data exploration analysis (DEA) and data cleaning before utilising it.

3.1 Processing of Traditional Feature Selection Models

The cardiovascular disease (CVD) data set was considered since we looked at feature selection methodology for choosing illness-associated characteristics from the human disease data set. Predicting the development of cardiovascular diseases (CVD) involves using machine learning classifiers that perform well with data and correctly identify individuals. Selecting the best course of action is still challenging even if machine learning (ML) algorithms can easily handle most categorisation issues. Numerous industrial challenges, such as required operation and analysis in production systems, are the focus of the data-driven solutions. Nevertheless, the most crucial strategy to employ with the ensemble is gathering difficulty-specific information. Using datasets, our model evaluated their learning capacity and used these ML classifiers to predict the incidence of cardiovascular illnesses (CVD). Building a more efficient classification model by selecting the most suitable machine learning classifier for the job and appropriately adjusting the hyperparameters can lead to accurate classification results. The machine learning algorithm receives the pertinent classes from the dataset.

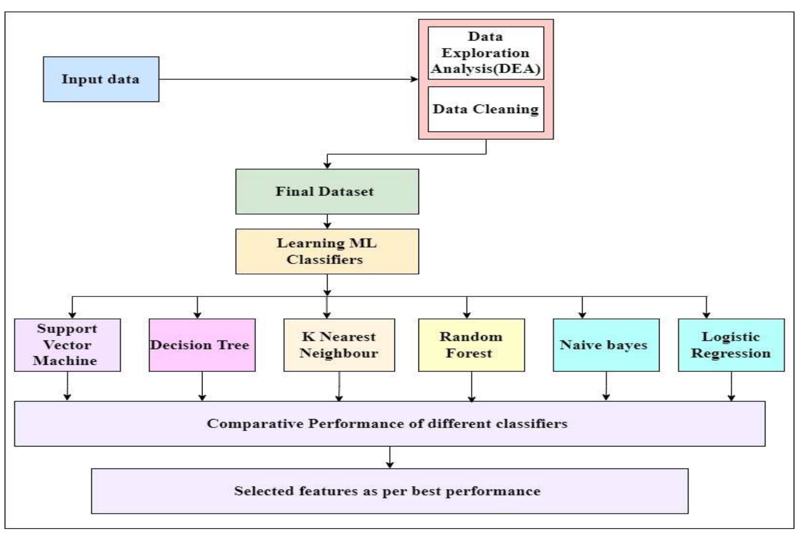


Figure 1: Individual method-based ML model

The data set used in this paper comprises 1025 patient records. The data set was then divided into training and testing records. Based on the produced conflict matrix, the latter was used to assess how well the former was presented. Previous models are essential for developing fixed procedures. Examining the confusion matrix allows for the analysis of several metrics and the evaluation of the classification models' efficacy.

Evaluation metrics can quantify the performance and efficacy of a statistical or machine-learning model. We may compare many models or algorithms with these measurements and see how well each performs. Precision and recall focus on the strength of the model's pessimistic and optimistic predictions, respectively, whereas accuracy assesses how comprehensive the model's predictions are. F1 Score is a comprehensive metric that evaluates classification models by finding an optimal balance between recall and accuracy. For the suggested model, we have considered the following machine-learning classifier techniques.

(a) K-NN Algorithm: We have used Classification and regression models using K-NN algorithm. It considers the new data point's nearest neighbour when estimating its value.
(b) Naive Bayes Algorithm: The Naive Bayes Theorem and the Bayes Theorem are based on the same principle. Naive Bayes Classifiers describe this family of classification techniques. It ranks high among the most effective and straightforward categorization methods.
(c) Logistic Regression: Logistic regression is a method that may be used for supervised learning. Its principal use is in classification tasks. The main goal of the model is to predict the probability that an instance will be a member of a specific class or not. Interactions between the set of independent variables and dependent binary variables are described.
(d) Random Forest Classifier: Supervised learning is the algorithm's method. When both regression and classification are involved, it is used. The process consists in combining several classifiers to tackle a complex problem.
(e) Decision Tree Classifier: Supervised learning is the algorithm's method. When both regression and classification are involved, it is used. Like a flowchart, it shows an algorithm's structure as a tree, with features represented by internal nodes, rules by branches, and the algorithm's output represented by leaf nodes.
(f) Support Vector Machine: A supervised algorithm is what this one is. Both classification and regression problems may be solved with it. Classification problems are readily solved with it.

3.2 Proposed ensemble model for feature selection

In the proposed model, the ensemble model is considered to determine the performance of CVD. Different methodologies (such as Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), and Support vector machine (SVM)) are considered to make ensemble models such as SVM+RF, DT+RF, LR+RF. After creating an ensemble method, the evaluation demonstrated accuracy and found better accuracy for the appropriate model. Once one model is fixed, CVD will be detected from the data.

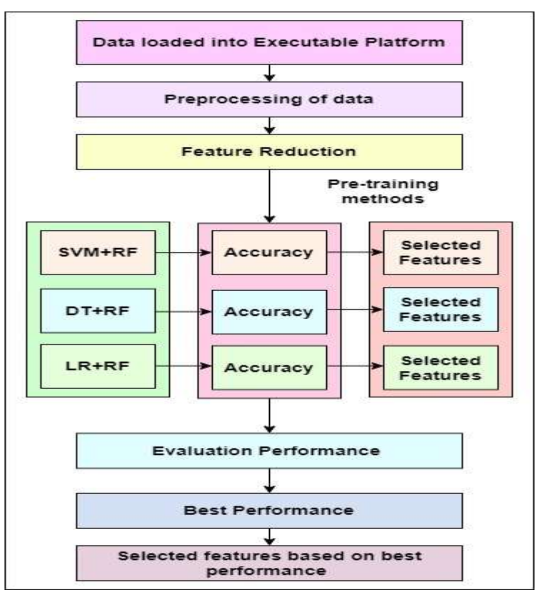


Figure 2: Ensemble methods-based ML Model

The whole process is mentioned in Figure 2. The model can be used for enhanced CVD with fewer properties and experts in health care services with incomplete clinical professionals for CVD. It can also serve as a proactive tool for CVD identification.

4. Experimental Results

4.1 Data set

For our suggested model, we have considered the Cardiovascular Disease Data Set [20]. We utilised the "Cardiovascular Disease Data set" as my source to verify the superior accuracy. There are one thousand records in the entire collection. Thirteen factors make up the data set: old peak, slope, roof main vessels, resting electron, exercise angina, goal, age, gender, chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, and so on. Information about categories and numbers are both included in the data collection. Furthermore, not a single value is empty. No duplicate patient ID is present.

4.2 Experimental Setup

The proposed model was executed using Python programming language with various packages. All experiments use machine learning classifiers and run with Windows-10 operating system. The hardware configuration is considered with Intel CoreTM i7-7700 HQ CPU@2.80GHz processor, Memory (RAM): 16.0GB, GeForce GTX 1060 GPU with 6GB GDDR5 memory.

4.3 Evaluation Metrics

We considered the Confusion matrices helpful tools for machine learning when dealing with classification problems. Our model's efficacy in action in this table summarises the number of cases correctly and wrongly identified for each category. We have considered different evaluation parameters, which are explained below.

(a) Accuracy: The frequency with which a model produces accurate predictions is called its accuracy. Despite its popularity, we shouldn't depend entirely on it unless we know its limits and strengths.

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

(b) Precision: Precision is a metric that expresses how near two measurements are to one another, either through measurement precision or consistency across time. It displays the level of precision or polish in a measuring procedure.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

(c) Recall: Recall is another critical metric for evaluating the performance of machine learning classification models. The main emphasis of recall is how well the model captures all the actual cases of success.

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

Evaluation metrics can quantify the performance and efficacy of a statistical or machine-learning model. You may compare many models or algorithms with these measurements and see how well each performs. Precision and recall focus on the strength of the model's pessimistic and optimistic predictions, respectively, whereas accuracy assesses how comprehensive the model's predictions are. F1 Score is a comprehensive metric that evaluates classification models by finding an optimal balance between recall and accuracy. As a result, we looked at several evaluation matrices according to the model shown in Fig. 3.

4.4 Performance on the proposed model

Using a variety of classifiers were constructed, and their performance was assessed to investigate their potential for forecast the occurrence of cardiovascular diseases (CVD). For evaluating our model, We looked at machine learning classifiers like Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support vector machine (SVM) K Nearest Neighbour (KNN), Naive Bayes (NB) . The dataset of the individual's undergoing evaluation was utilized to detect cardiovascular diseases (CVD). The categorization models described here aim to classify each individual as normal or aberrant (PD). The training dataset comprises 75% pre-classified data, while the test dataset comprises 25% unclassified data points. Machine algorithms iteratively improve the classification performance during model construction by utilizing a training dataset. The evaluation metrics using classifiers are mentioned in Figure 3.

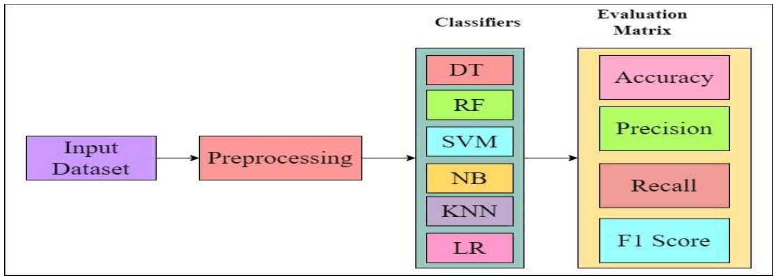


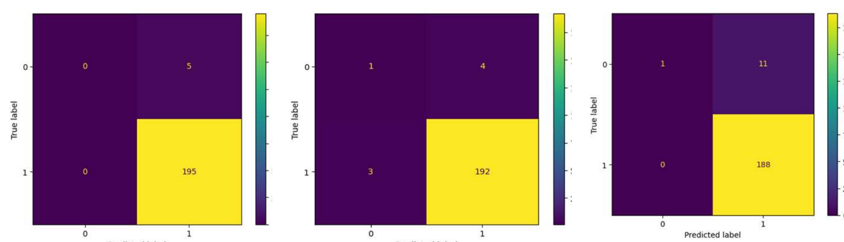
Figure 3: Evaluation matrices of ML Classifiers

The fixed model is trained by examining the trained record. Each data point in this dataset is only characterized by the quality of the data. The confusion matrix contains four types: TP, TN, FP, and FN. The confusion matrix presents the ML classification algorithm, which derives several performance measures. The percentage of test samples categorised as TP, TN, FP, or FN is shown in the confusion matrices. Several measures were used to evaluate the classifier's performance, including accuracy, precision, recall, and sensitivity. The evaluation's performance and category are displayed as shown in Figure (5-9). We also looked at several machine learning classifiers, including SVM, DT, LR, and RF, along with the hybrid model that goes with it such as SVM+RF, DT+RF, LR+RF.

We applied SVM, DT, LR, and RF classifiers on the appropriately trained and fit-ted classification models to consider the confusion matrices. After then, these models were put to the test. While TN and FP were correctly classified by the fitted SVM-based categorization model, TP and FN were not. Similarly, it is possible to verify the accuracy of additional classifiers. The SVM classifier excels in specificity and precision, whereas the RF classifier performs better than all other traditional methods regarding sensitivity and accuracy. Furthermore, while the SVM classifier performs a decent task at detecting CVD, it isn't as sensitive or accurate at identifying those with the rule. It lacks specificity and precision yet yields exact results.

4.4.1 Evaluation Metrics Performance

As per the evaluation metrics, we have considered confusion metrics, as shown in Figure 4. We have considered ensemble methods for confusion metrics combined with two performance methods.



SVM+RF **DT+RF** **LR+RF**

Figure 4: Confusion matrix on ensemble model

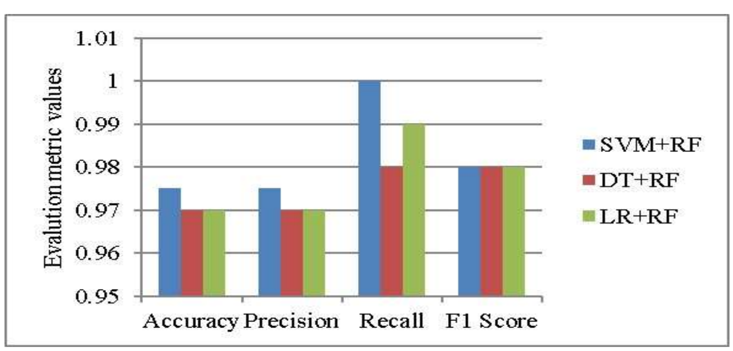


Figure 5: Evaluation metrics performance on ensemble model

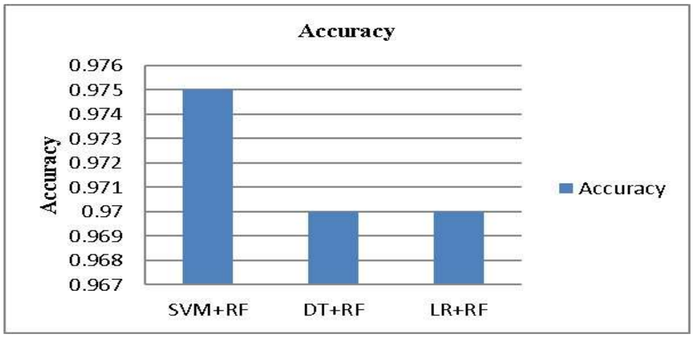


Figure 6: Accuracy on ensemble model

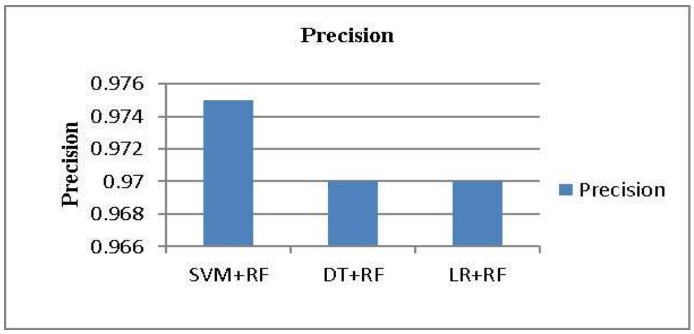


Figure 7: Precision on ensemble model

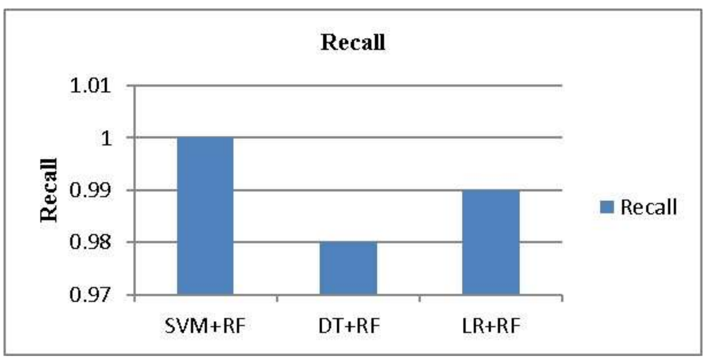


Figure 8: Recall on ensemble model

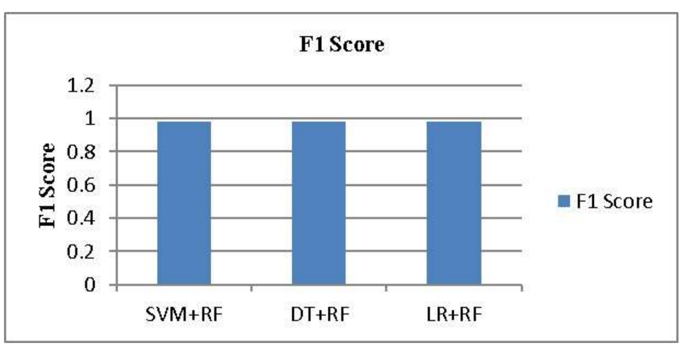


Figure 9: F1 Score on ensemble model

4.5 Comparative performance analysis

We have considered the comparative performance analysis with an existing and proposed method. We observed that most of the accuracy of the existing method is less than the proposed ensemble methods. The ensemble performed well (with more than 95 % accuracy). Our proposed model performed well in accuracy and other evaluation metrics such as precision, recall, F1-Score, etc. with more than 90% score in each evaluation item as shown in figure 5.

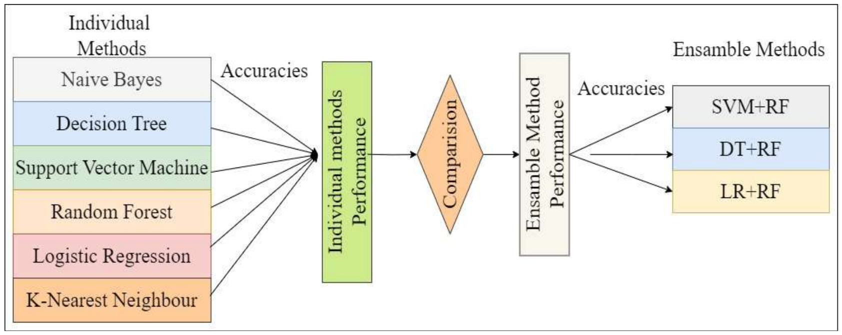


Figure 10: Comparative performance between Individual Methods and our ensemble methods

Since our proposed model is better than the existing model, we considered the final ensemble model based on the better performance of the method i.e LR+RF model, which performs better than other methods, as shown in Figure 10. We have compared our evaluation performance with [23], and we have a better performance than the existing performance, as shown in Table 1.

Table 1: Comparative accuracy performance with [23]

S.No.	Existing Methods	Accuracy	Proposed Methods	Accuracy
1	NB, BN, RF, and MLP	85.48	LR and RF	97.0
2	RBF and SVM	92.22	SVM and RF	97.5
3	Randomized decision tree ensemble	93	DT and RF	97.0

Table 1 shows that the existing ensemble model performs less than the proposed model, as per [23]. For example, the existing model based on (NB, BN, RF, and MLP) has accuracy (85.48%) whereas the proposed model based on (LR and RF) has 97.0%. Similarly, the existing model based on (RBF and SVM) has accuracy (92.22) whereas our proposed model based on (SVM and RF) has accuracy (97.5%). Thus, our proposed model performed better than the existing model.

4.6 Result analysis

We have considered data for processing through a generic model, as shown in Fig 3, where different classifiers are performed with their accuracy. Figure 11 indicates the correlation coefficient for the attributes in the "Cardiovascular Disease Data set".

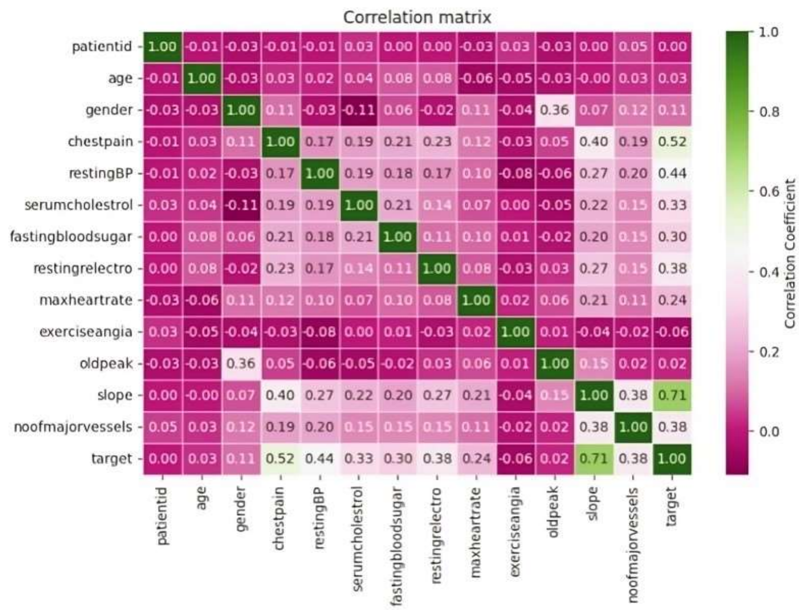


Figure 11: Correlation coefficient among features of Cardiovascular_Disease_Data set

Table 2 represents the Accuracy, Precision and Recall values on the total data set with 12 features.

Table 2: Evaluation of performance on traditional methods

Model	Accuracy	Precision	Recall
DT	0.965	0.958	0.982
RF	**0.980**	0.982	0.982
SVM	0.585	**0.983**	**0.991**
NB	0.81	0.796	0.905
K-NN	0.515	0.973	0.948
LR	0.585	0.966	0.974
ANN	0.956	0.940	0.981

The following picture represents the accuracy of various ML models on the total data set with 12 features.

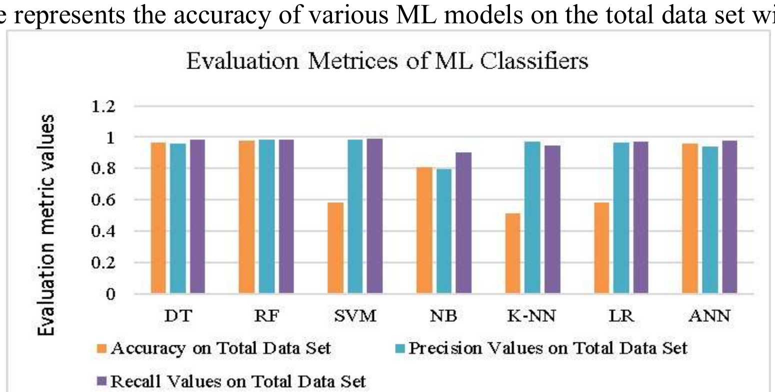


Fig.12: Accuracy, Precision and Recall of LR, DT, RF, SVM, NB, K- NN, ANN on complete data set

In all these accuracies, Random Forest had the highest accuracy, and the K-NN had the lowest, as shown in Figure 12. Similarly, the optimal performance of each parameter is considered in Table 2 and identified with bold values.

4.7 Selected features as per model

Feature selection method is used to build predictive models using the fewest possible input variables. By analysing the statistical link between each input variable and the objective variable, statistical feature selection algorithms detect which input variables strongly correlate with the target variable. As per the dataset, we have considered two types of feature set such as (a) numeric type feature: {"age", "restingBP", "serumcholestrol", "maxheartrate", "oldpeak"}, and (b) Categorical type of feature: {"gender", "chestpain", "fastingbloodsugar", "restingrelectro", "exerciseangia", "slope", "noofmajorvessels", "target"}. Asper the proposed model, we have selected the features such as: {patientid', 'age', 'gender', 'chestpain', 'restingBP', 'serumcholestrol'}. Above features are predicted for the cardiovascular disease as per the model.

5. Conclusion and Future Work

We have considered the development of predictive power for several machine learning ensemble models to evaluate the cardiovascular disease. The Random Forest method is considered to determine the best machine learning model, and combined with another method to make an ensemble model. Our ensemble model is based on three approaches such as SVM+RF, DT+RF, LR+RF. When we demonstrated our model with CVD data set, we found our model provided better results than traditional model. Recall and accuracy were both at satisfactory levels for both models. The ensemble method on Support Vector Machine and Random Forest holds excellent potential for estimating the risk of cardiovascular disease. This ensemble method selects effective features with better performance.

Consequently, our research paves the way for the creation of potent preventative measures against cardiovascular disease. More investigation into other datasets and machine learning techniques is required to validate its value, which may produce better outcomes. As a result, this paper suggested an ensemble model to find appropriate features from patient data to create correct machine-learning models for predicting cardiovascular disease. We have a strategy to create a flexible ensemble model and evaluate model predictions using AI in the future.

References

[1] A. Hafiz and N. Kaur, “Heart Disease Prediction based on Machine learning Technique,” 2023. doi: 10.1109/TEMSMET56707.2023.10150023.
[2] Bhuyan H. K., Vinayakumar Ravi, Biswajit Brahma, Nilayam Kumar Kamila, Disease analysis using machine learning approaches in healthcare system, Health and Technology, Vol. 12, Issue-5, pages: 987-1005, 2022.
[3] Madhumita Pal, Smita Parija, corresponding author Ganapati Panda, Kuldeep Dhama, and Ranjan K. Mohapatra, Risk prediction of cardiovascular disease using machine learning classifiers, Open Med (Wars). 17(1): 1100–1113, 2022.
[4] Bhuyan H. K., Vinay Kumar Ravi, Analysis of Sub-feature for Classification in Data Mining, IEEE Transaction on Engineering Management, Volume: 70, Issue: 8, Page(s): 2732-2746, 2023.
[5] Bhuyan H. K., Narendra Kumar Kamila, Privacy preserving sub-feature selection based on fuzzy probabilities, Cluster computing, Vol-17, Issue-4, PP. 1383-1399, 2014. DOI https://doi.org/10.1007/s10586-014-0393-9
[6] Ghulam Ali; Aqsa Dastgir; Muhammad Waseem Iqbal; Muhammad Anwar; Muhammad Faheem, A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification From Fundus Images, IEEE Journal of Translational Engineering in Health and Medicine (Volume: 11), Page(s): 341 - 350, 2023.
[7] Mohammed B. Abubaker; Bilal Babayiğit, Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods, IEEE Transactions on Artificial Intelligence, Volume: 4, Issue: 2, April 2023.
[8] K. Bagadi *et al.*, “Cardiovascular Disease Prediction Using Machine Learning Algorithms,” in *IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, Oct. 2023, pp. 1–8. doi: 10.1109/ICETAS59148.2023.10346353.
[9] C. Sushama, P. Arulprakash, M. Sunil Kumar, D. Ganesh, and K. Sujatha, “The Future of Education: Artificial Intelligence based Remote Learning,” *Int. J. Early Child. Spec. Educ. (INT-JECSE*, vol. 14, no. 03, p. 2022, 2022, doi: DOI: 10.9756/INT-JECSE/V14I3.478.
[10] Subhash Mondal, Ranjan Maity, Yachang Omo, Soumadip Ghosh, Amitava Nag, An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches, IEEE Access, VOLUME 12, Page(s): 7255 - 7270, 2024.
[11] Aqsa Rahim; Yawar Rasheed; Farooque Azam; Muhammad Waseem Anwar; Muhammad Abdul Rahim, Abdul Wahab Muzaffar, An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases, IEEE Access (Volume: 9), Page(s): 106575 - 106588, 2021.

[12] A. Azarudeen, P. Hemavathy, A. M. Jabbar, S. J. Abed, F. Fhadil, and A. Ali, "Forecasting of Cardiovascular Disease Using Machine Learning," in *3rd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2023*, 2023, pp. 179–182. doi: 10.1109/ICACITE57410.2023.10183251.
[13] A. M. Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning," *IEEE Access*, vol. 11, pp. 56214–56224, 2023, doi: 10.1109/ACCESS.2023.3281484.
[14] K. Karthik, A. L. Reddy, R. Kulkarni, and M. J. Mehdi, "Algorithm Accuracy Verification in Heart Disease Analysis using Machine Learning," in *Proceedings of the 2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023*, 2023, pp. 345–349. doi: 10.1109/ICAAIC56838.2023.10140446.
[15] A. D. Dhruva, P. B., S. Kamepalli, S. S. S, and S. Kunisetti, "An efficient mechanism using IoT and wireless communication for smart farming," *Mater. Today Proc.*, pp. 1–6, 2023, doi: 10.1016/j.matpr.2021.07.363.
[16] M. Gagoriya and M. K. Khandelwal, "Heart Disease Prediction Analysis Using Hybrid Machine Learning Approach," in *Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2023*, 2023, pp. 896–899. doi: 10.1109/IITCEE57236.2023.10090896.
[17] Bhuyan H. K., Vinay Kumar Ravi, An Integrated Framework with Deep learning for Segmentation and Classification of Cancer Disease, Int J. on Artificial Intelligence Tools (IJAIT), Vol. 32, No. 02, 2340002, 2023.
[18] Bhuyan H. K.,A.Vijayaraj, Vinay Kumar Ravi, Development of Secrete Images in Image Transferring System, Multimedia Tools and Applications 82 (5), 7529-7552, 2023.
[19] K. Damodar Prabhu, P. Rao, K. Varsha Bhat, N. S. Pooja, and P. R. Kamath, "Detection and Analysis of Cardiovascular Diseases using Machine Learning Techniques," in *IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, DISCOVER 2023*, 2023, pp. 258–262. doi: 10.1109/DISCOVER58830.2023.10316703.
[20] J. Zheng et al., "Optimal multi-stage arrhythmia classification approach," Sci. Rep., vol. 10, no. 1, 2020, Art. no. 2898.
[21] Bhuyan H. K., M Saikiran, Murchhana Tripathy, Vinayakumar Ravi, Wide-ranging approach-based feature selection for classification, Multimedia Tools and Applications, pages: Volume 82, Issue 15, pp 23277–23304, Jun 2023. https://doi.org/10.1007/s11042-022-14132-z
[22] Bhuyan H. K., A Vijayaraj, Vinayakumar Ravi, Diagnosis system for cancer disease using a single setting approach, Multimedia Tools and Applications, Springer US, pp. 1-27, 2023.
[23] Ibomoiye Domor Mienye, Yanxia Sun, Zenghui Wang, An improved ensemble learning approach for the prediction of heart disease risk, Informatics in Medicine Unlocked, Volume 20, 100402, 2020.
[24] Hemanta Kumar Bhuyan, L Raghu Kumar, K Ramakrishna Reddy, Optimization model for Sub-feature selection in data mining, 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE Explore, Pages: 1212-1216, 2019.
[25] Bhuyan H. K, CV Madhusudan Reddy, Sub-feature selection for novel classification, 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE Explore, Pages: 477-482, 2018.