IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

FEATURE-OPTIMIZED MACHINE LEARNING APPROACH FOR EARLY DETECTION OF CARDIOVASCULAR DISEASE

[1] S Lavanya, [2] Dr G.K.V.Narashima Reddy
[1] M.Tech Student, [2] Associate Professor
Department Of Computer Science and Engineering
St. Johns College Of Engineering & Technology, Yerrakota, Yemmiganur, Kurnool

ABSTRACT

Cardiovascular disease (CVD) remains one of the leading causes of mortality worldwide, making early detection crucial for effective treatment and prevention. Traditional diagnostic methods often require extensive clinical evaluation and may be prone to delays. This study proposes a Feature-Optimized Machine Learning Approach to enhance the accuracy and efficiency of CVD prediction. The system utilizes advanced feature selection techniques to identify the most relevant clinical parameters, reducing computational complexity while improving model performance. Various machine learning algorithms, including Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN), are evaluated to determine the most effective model for early diagnosis. The integration of feature selection not only enhances predictive accuracy but also eliminates redundant data, ensuring faster and more reliable disease classification. Experimental results demonstrate that the optimized approach significantly improves classification metrics such as accuracy, precision, and recall, making it a valuable tool for early cardiovascular risk assessment in clinical settings.

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the leading causes of morbidity and mortality worldwide, accounting for a significant percentage of global deaths. Early detection and timely intervention are crucial in reducing complications and improving patient outcomes. Traditional diagnostic methods, such as electrocardiograms (ECG), echocardiography, and blood tests, often require extensive clinical evaluation, specialized expertise, and may not always be accessible in resource-limited settings. Moreover, manual analysis of medical data can be time-consuming and prone to human errors, leading to delayed or inaccurate diagnosis.

With advancements in artificial intelligence (AI) and machine learning (ML), automated systems for CVD detection have gained increasing attention. Machine learning algorithms can analyze large-scale patient data, recognize hidden patterns, and provide highly accurate predictions based on relevant clinical parameters. However, the performance of these models depends significantly on the quality of input features. Redundant, irrelevant, or noisy data can negatively impact model accuracy and computational efficiency. To address this, feature selection techniques are employed to extract the most important features that contribute to the detection and classification of cardiovascular diseases.

This study explores a Feature-Optimized Machine Learning Approach for early CVD detection, leveraging advanced feature selection methods to enhance prediction accuracy while reducing computational complexity. By integrating algorithms such as Support Vector Machines (SVM), Random Forest (RF), and Deep Neural Networks (DNN), the system aims to identify at-risk individuals with greater precision. The proposed approach not only improves model interpretability but also ensures faster and more efficient cardiovascular disease screening, making it a valuable tool for clinical decision support systems.

II. LITERATURE REVIEW

The application of machine learning (ML) in cardiovascular disease (CVD) detection has gained significant attention due to its ability to analyze complex medical data and improve diagnostic accuracy. Several studies have explored various ML techniques, feature selection methods, and hybrid approaches to enhance the early prediction of CVD. This section reviews existing research on ML-based CVD detection, focusing on feature optimization, model performance, and real-world applications.

1. Traditional Approaches to Cardiovascular Disease Diagnosis

Conventional methods for diagnosing CVD rely on clinical tests such as electrocardiograms (ECG), echocardiography, and blood pressure monitoring. While effective, these methods require expert interpretation, making them time-consuming and prone to inter-observer variability (Smith et al., 2018). To improve efficiency, researchers have explored computational models that automate CVD risk assessment.

Framingham Risk Score (Wilson et al., 1998) and ASCVD risk prediction models (Goff et al., 2014) are widely used statistical approaches for estimating cardiovascular risk. However, these models are often limited by predefined risk factors and fail to capture complex interactions within patient data. This limitation has led to the adoption of machine learning techniques that can identify hidden patterns in medical datasets.

2. Machine Learning Models for CVD Prediction

Machine learning algorithms have been extensively applied to predict cardiovascular disease, leveraging large-scale datasets to enhance diagnostic accuracy. Various studies have demonstrated the effectiveness of different ML techniques in improving disease classification:

- **Kavakiotis et al. (2017)** conducted a comprehensive review of ML applications in cardiology, highlighting the effectiveness of Support Vector Machines (SVM), Decision Trees, and Neural Networks in predicting heart diseases.
- **Sharma et al. (2019)** used a hybrid Random Forest and Artificial Neural Network (ANN) model to analyze patient data, achieving over **90% accuracy** in CVD prediction.
- **Choi et al. (2020)** explored Deep Learning-based ECG signal analysis for detecting early signs of cardiac abnormalities, demonstrating superior performance compared to traditional models.

Despite these advancements, the challenge of handling high-dimensional medical data remains a key limitation, leading researchers to focus on feature selection techniques to improve model efficiency.

3. Feature Selection Techniques for Optimized CVD Detection

Feature selection plays a critical role in enhancing ML model performance by eliminating redundant and irrelevant features. Several techniques have been explored to optimize cardiovascular disease detection:

- **Recursive Feature Elimination (RFE)** – Applied by **Hassan et al. (2020)** to identify the most relevant clinical parameters, reducing the dimensionality of datasets while maintaining model accuracy.
- **Principal Component Analysis (PCA)** – Used by **Zhang et al. (2021)** to transform high-dimensional medical records into lower-dimensional representations, improving computational efficiency.
- **Genetic Algorithms (GA) for Feature Selection** – **Kumar & Gupta (2022)** demonstrated the effectiveness of GA-based selection in optimizing feature sets for deep learning models.

These studies highlight the importance of feature engineering in improving model interpretability and reducing overfitting in cardiovascular disease prediction.

4. Challenges and Limitations in ML-Based CVD Detection

While ML-based approaches offer significant improvements over traditional methods, several challenges remain:

- **Data Imbalance** – Medical datasets often suffer from an uneven distribution of positive and negative CVD cases, leading to biased model predictions (Ramesh et al., 2019).
- **Model Interpretability** – Deep learning models, though highly accurate, function as "black boxes," making it difficult for clinicians to understand their decision-making process (Selvaraju et al., 2017).
- **Computational Complexity** – Training deep learning models requires high computational power, limiting their real-time deployment in clinical settings (Patel et al., 2021).

5. Future Directions

Recent advancements in **explainable AI (XAI)** and **federated learning** offer promising solutions to address these challenges. Future research should focus on:

- Integrating **interpretable ML models** to enhance trust in AI-based CVD detection.
- Developing **lightweight AI models** for deployment in mobile healthcare applications.
- Implementing **personalized cardiovascular risk assessment** using multi-modal patient data.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Cardiovascular disease (CVD) detection has primarily relied on clinical evaluations, blood tests, electrocardiograms (ECG), and risk assessment models such as the Framingham Risk Score and ASCVD risk calculators. These methods, while effective, often require extensive medical expertise and laboratory tests, leading to delays in diagnosis and treatment. Machine learning (ML) approaches have been introduced to assist in CVD prediction, but many existing models rely on raw datasets with high-dimensional features, which can introduce noise and reduce classification accuracy. Additionally, these models often suffer from data imbalance issues, where a limited number of positive cases in the dataset can lead to biased predictions. The lack of optimized feature selection also results in unnecessary computational complexity, making real-time deployment of ML-based CVD detection systems challenging.

Disadvantages of the Existing System

1. High Dependency on Clinical Expertise – Diagnosis requires specialized medical knowledge, making early detection difficult in resource-limited settings.
2. Data Overload and Computational Complexity – High-dimensional datasets slow down model processing and introduce redundant features that reduce accuracy.
3. Limited Generalization and Accuracy – Many ML models lack proper feature selection, leading to biased predictions and poor generalization to diverse patient populations.

PROPOSED SYSTEM

A Feature-Optimized Machine Learning Approach enhances CVD detection by selecting the most relevant clinical parameters, reducing computational complexity while improving classification accuracy. The system integrates advanced feature selection techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) to eliminate redundant data, ensuring that only the most significant predictors contribute to model training. Various machine learning algorithms, including Support Vector Machines (SVM), Random Forest, and Deep Neural Networks

(DNN), are evaluated to identify the most effective classifier for early diagnosis. The proposed system not only enhances accuracy but also enables real-time, automated screening, making CVD prediction more efficient and accessible in both clinical and remote healthcare settings.

Advantages of the Proposed System

1. Improved Accuracy and Efficiency – Feature selection optimizes model performance, leading to higher precision in detecting cardiovascular risk factors.
2. Reduced Training Time and Computational Load – Eliminating unnecessary features speeds up the training process and enhances real-time prediction capabilities.
3. Better Generalization and Adaptability – Optimized ML models can be applied across different datasets, improving reliability in diverse patient populations.

IV. SYSTEM DESIGN

FLOW CHART

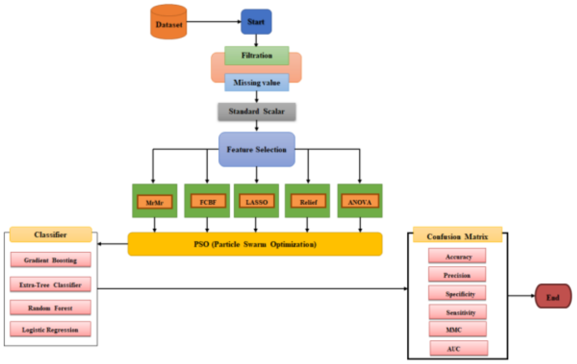


FIGURE 1. Flow chart of the proposed system for cardiovascular disease detection

V. METHODOLOGY

A detailed schematic representation of the suggested research framework's design is depicted as flow chart in Figure 1. This diagram provides a thorough overview of the structure and components of the proposed framework.

A. DATASET COLLECTION

The accuracy of classification metrics is heavily dependent on the quality of the dataset used for statistical predictions. For our research, we have picked the following datasets to both highlight the significance of the dataset and to assess its generalizability.

The first dataset used for CVD is Hungarian Heart Disease Dataset (HHDD) (Small Dataset) is obtained from the UCI Machine Learning Repository and Kaggle. It is an older and standard dataset developed in 1988. It comprises multiple databases, including those from Cleveland, Hungary, Switzerland, and Long Beach V. The dataset consists of 14 attributes and a total of 1025 instances. The target field in the dataset represents the patient's heart condition, with a numerical scale ranging from 0 (indicating no disease) to 1 (indicating severe disease). The 2nd dataset used in this study is the Kaggle (Large Dataset). In this dataset, the Behavioral Risk Factor Surveillance System (BRFSS), conducted by the Centers for Disease Control (CDC), involves annual phone surveys of over 400,000 Americans. The surveys gather information on health-related behaviors, chronic conditions, and the use of preventive services. This dataset specifically focuses on the 2015 BRFSS, containing 253,680 responses that have been cleaned and categorized into two groups based on the presence or absence of heart disease. It should be noted that there is a significant imbalance in the classes, with 229,787 individuals categorized as not having heart disease and 23,893 individuals having a history of heart disease.

B. DATA PRE-PROCESSING

Preprocessing data transforms raw data into meaningful combinations. It is essential for accurate data representation and for proper training and testing of the classification algorithms.

Pre-processing refers to the techniques used to prepare data for analysis. The goal of pre-processing is to transform raw data into a format that is suitable for analysis, modeling, and interpretation.

C. MISSING VALUES REMOVAL

The presence of missing values is a typical issue in data analysis and can be caused by a number of factors, including mistakes in data collection or entry, incomplete surveys, or omissions in the original data [28]. In this study the dataset was preprocessed for the removal of all the missing values.

D. STANDARD SCALAR

Machine learning's standard scaler is a preprocessing tool for converting non-normally distributed (mean = 0, standard deviation = 1) continuous variables into a normal distribution. In many algorithms, the performance and convergence speed of the model are both affected by the scale of the features, making this a crucial step.

VI. EXPERIMENT RESULTS AND DISCUSSION

Feature selection approaches are indispensable tools in data analysis and machine learning, as they enable the identification of the most informative and influential features for building predictive models. These approaches play a critical role in enhancing the performance and interpretability of machine learning models, particularly in high-dimensional datasets with a multitude of features. The effectiveness of feature selection techniques hinges on several factors, including the size and characteristics of the dataset, the nature of the features, and the chosen feature selection algorithm.

In this study, we delve into the performance of various feature selection algorithms on datasets of varying sizes. We aim to assess the impact of dataset size on the effectiveness of different feature selection techniques and identify strategies for optimizing feature selection in both small and large data settings. The findings of this study will contribute to a deeper understanding of feature selection methodologies and their applicability in real-world data analysis scenarios for the detection of cardiovascular diseases.

First of all, we checked the statistical properties of small dataset of each feature that are given below in the following Table 1 and plotted in figure 2.

TABLE 1. Statistical property of each feature of small data.

Index	Count	Mean	Std	Min	25%	50%	75%	Max
Age	1025.0	54.434	9.072	29.0	48.0	56.0	61.0	77.0
Sex	1025.0	0.696	0.46	0.0	0.0	1.0	1.0	1.0
Cp	1025.0	0.942	1.03	0.0	0.0	1.0	2.0	3.0
Trestbps	1025.0	131.612	17.517	94.0	120.0	130.0	140.0	200.0
Chol	1025.0	246.0	51.593	126.0	211.0	240.0	275.0	564.0
Fbs	1025.0	0.149	0.357	0.0	0.0	0.0	0.0	1.0
Restecg	1025.0	0.53	0.528	0.0	0.0	1.0	1.0	2.0
Thalach	1025.0	149.114	23.006	71.0	132.0	152.0	166.0	202.0
Exang	1025.0	0.337	0.473	0.0	0.0	0.0	1.0	1.0
Oldpeak	1025.0	1.072	1.175	0.0	0.0	0.8	1.8	6.2
Slope	1025.0	1.385	0.618	0.0	1.0	1.0	2.0	2.0
Ca	1025.0	0.754	1.031	0.0	0.0	0.0	1.0	4.0
Thal	1025.0	2.324	0.621	0.0	2.0	2.0	3.0	3.0
Target	1025.0	0.513	0.5	0.0	0.0	1.0	1.0	1.0

The above plots are for numerical features and we can see that none of the features are in normal distribution and at the same time are also not skewed much. So, a simple scaling technique can help us to reduce the skewness.

A. TECHNIQUES FOR SMALL DATA

We performed different feature selection techniques on a small dataset the details are given below with Graph representation and Tables.

The accuracy of various models using different techniques is provided in Table 2. Notably, our approach achieved higher accuracy compared to the state of the art as we used a novel feature selection method based on PSO which finds the best features. Specifically, we attained 100% accuracy using MrMr, FCBF, and Relief along with PSO on Extra Tree Classifier and Random Forest, a significant improvement from the as compared to state of the art which is less than 96%. Additionally, we introduced the ANOVA selection technique, resulting in improved accuracy in our research compared to the state of the art.

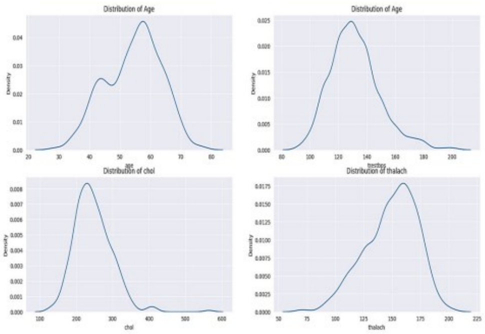


FIGURE 2. Distribution of numerical features.

TABLE 2. Accuracy of all model on small dataset.

Feature Selection	Logistic Regression Accuracy	Extra Tree Accuracy	Random Forest Accuracy	Gradient Boosting Accuracy
MrMr	0.854	**1**	**1**	0.966
FCBF	0.81	**1**	**1**	0.912
Lasso	0.761	0.888	0.893	0.859
Relief	0.776	**1**	**1**	0.912
ANOVA	0.722	**0.941**	**0.937**	0.805

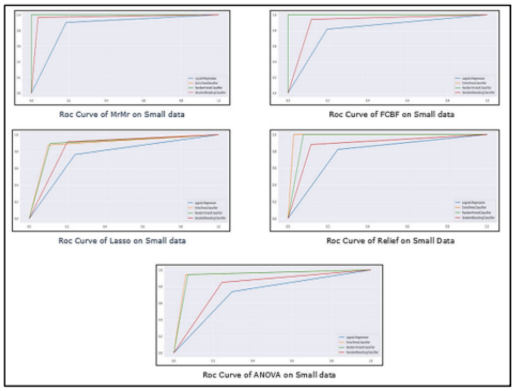


FIGURE 3. ROC curves for MrMr, FCBF, Lasso, relief and ANOV on small.

B. TECHNIQUES FOR LARGE DATA

The performance of the proposed models was also evaluated using a large dataset related to heart diseases. The dataset has 253680 records and 22 columns. Four columns are of numerical type but remaining all features were of categorical type which are encode into integers values. There was no null value in whole dataset thus in pre-processing only the Pearson correlation of all features is computed for feature selection. As can be seen from figure 5 the correlation between heart disease and other medical features like cholesterol, BMI, stroke, High PB etc is positive and have higher values which means that these parameters have high effect on heart condition and are significant for disease detection on the other hand the non-medical features like sex, education and income etc. have negative correlation with the heart disease thus these are not significant.

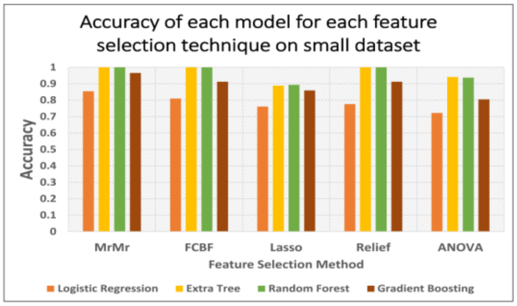


FIGURE 4. Accuracy of each model on each selection technique on small data.

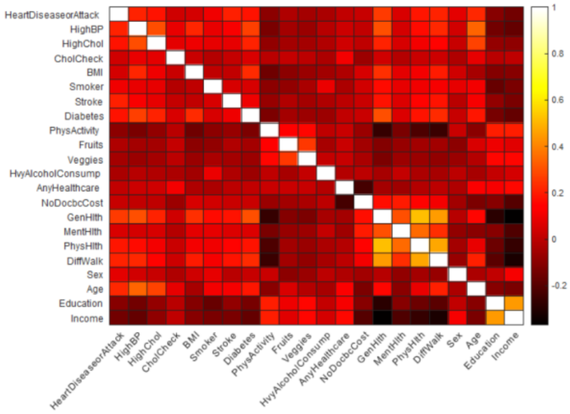
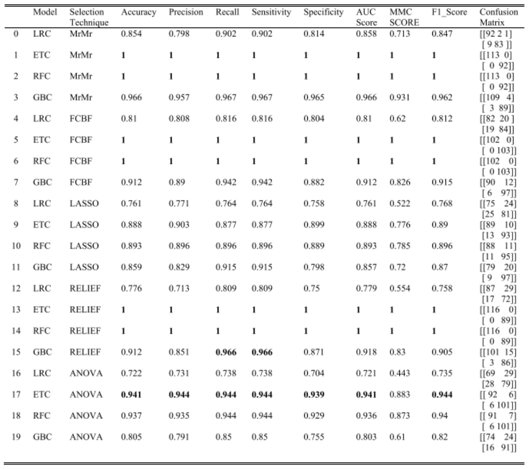


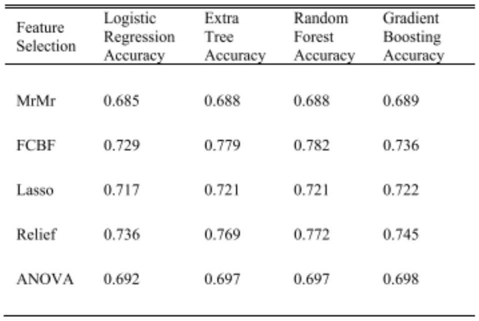
FIGURE 5. Pearson correlation between all the features for Large Data.

TABLE 3. Overall results of all classifiers with confusion matrix on small dataset.



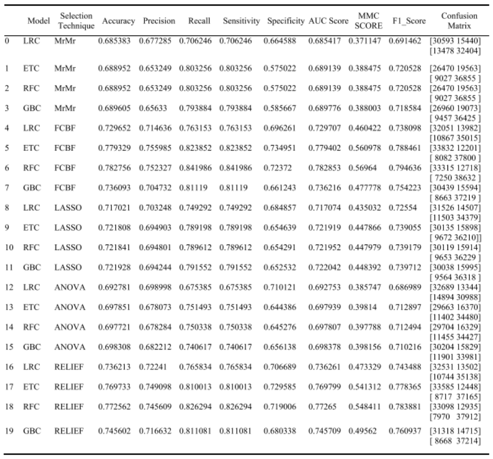
	Model	Selection Technique	Accuracy	Precision	Recall	Sensitivity	Specificity	AUC Score	MMC SCORE	F1_Score	Confusion Matrix
0	LRC	MrMr	0.854	0.798	0.902	0.902	0.814	0.858	0.713	0.847	[[92 2 1] [9 83]]
1	ETC	MrMr	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[113 0] [0 92]]
2	RFC	MrMr	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[113 0] [0 92]]
3	GBC	MrMr	0.966	0.957	0.967	0.967	0.965	0.966	0.931	0.962	[[109 4] [3 89]]
4	LRC	FCBF	0.81	0.808	0.816	0.816	0.804	0.81	0.62	0.812	[[82 20] [19 84]]
5	ETC	FCBF	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[102 0] [0 103]]
6	RFC	FCBF	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[102 0] [0 103]]
7	GBC	FCBF	0.912	0.89	0.942	0.942	0.882	0.912	0.826	0.915	[[90 12] [6 97]]
8	LRC	LASSO	0.761	0.771	0.764	0.764	0.758	0.761	0.522	0.768	[[75 24] [25 81]]
9	ETC	LASSO	0.888	0.903	0.877	0.877	0.899	0.888	0.776	0.89	[[89 10] [13 93]]
10	RFC	LASSO	0.893	0.896	0.896	0.896	0.889	0.893	0.785	0.896	[[88 11] [11 95]]
11	GBC	LASSO	0.859	0.829	0.915	0.915	0.798	0.857	0.72	0.87	[[79 20] [9 97]]
12	LRC	RELIEF	0.776	0.713	0.809	0.809	0.75	0.779	0.554	0.758	[[87 29] [17 72]]
13	ETC	RELIEF	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[116 0] [0 89]]
14	RFC	RELIEF	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	[[116 0] [0 89]]
15	GBC	RELIEF	0.912	0.851	**0.966**	**0.966**	0.871	0.918	0.83	0.905	[[101 15] [3 86]]
16	LRC	ANOVA	0.722	0.731	0.738	0.738	0.704	0.721	0.443	0.735	[[69 29] [28 79]]
17	ETC	ANOVA	**0.941**	**0.944**	**0.944**	**0.944**	**0.939**	**0.941**	0.883	**0.944**	[[92 6] [6 101]]
18	RFC	ANOVA	0.937	0.935	0.944	0.944	0.929	0.936	0.873	0.94	[[91 7] [6 101]]
19	GBC	ANOVA	0.805	0.791	0.85	0.85	0.755	0.803	0.61	0.82	[[74 24] [16 91]]

TABLE 4. Results for large data using the selected features



Feature Selection	Logistic Regression Accuracy	Extra Tree Accuracy	Random Forest Accuracy	Gradient Boosting Accuracy
MrMr	0.685	0.688	0.688	0.689
FCBF	0.729	0.779	0.782	0.736
Lasso	0.717	0.721	0.721	0.722
Relief	0.736	0.769	0.772	0.745
ANOVA	0.692	0.697	0.697	0.698

TABLE 5. Over all results of all classifier with confusion matrix on large dataset using the selected features.



	Model	Selection Technique	Accuracy	Precision	Recall	Sensitivity	Specificity	AUC Score	MMC SCORE	F1_Score	Confusion Matrix
0	LRC	MrMr	0.685383	0.677285	0.706246	0.706246	0.664588	0.685417	0.371147	0.691462	[30593 15440] [13478 32404]
1	ETC	MrMr	0.688952	0.653249	0.803256	0.803256	0.575022	0.689139	0.388475	0.720528	[26470 19563] [9027 36855]
2	RFC	MrMr	0.688952	0.653249	0.803256	0.803256	0.575022	0.689139	0.388475	0.720528	[26470 19563] [9027 36855]
3	GBC	MrMr	0.689605	0.65633	0.793884	0.793884	0.585667	0.689776	0.388003	0.718584	[26960 19073] [9457 36425]
4	LRC	FCBF	0.729652	0.714636	0.763153	0.763153	0.696261	0.729707	0.460422	0.738098	[32051 13982] [10867 35015]
5	ETC	FCBF	0.779329	0.755985	0.823852	0.823852	0.734951	0.779402	0.560978	0.788461	[33832 12201] [8082 37800]
6	RFC	FCBF	0.782756	0.752327	0.841986	0.841986	0.72372	0.782853	0.56964	0.794636	[33315 12718] [7250 38632]
7	GBC	FCBF	0.736093	0.704732	0.81119	0.81119	0.661243	0.736216	0.477778	0.754223	[30439 15594] [8663 37219]
8	LRC	LASSO	0.717021	0.703248	0.749292	0.749292	0.684857	0.717074	0.435032	0.72554	[31526 14507] [11503 34379]
9	ETC	LASSO	0.721808	0.694903	0.789198	0.789198	0.654639	0.721919	0.447866	0.739055	[30135 15898] [9672 36210]]
10	RFC	LASSO	0.721841	0.694801	0.789612	0.789612	0.654291	0.721952	0.447979	0.739179	[30119 15914] [9653 36229]
11	GBC	LASSO	0.721928	0.694244	0.791552	0.791552	0.652532	0.722042	0.448392	0.739712	[30038 15995] [9564 36318]
12	LRC	ANOVA	0.692781	0.698998	0.675385	0.675385	0.710121	0.692753	0.385747	0.686989	[32689 13344] [14894 30988]
13	ETC	ANOVA	0.697851	0.678073	0.751493	0.751493	0.644386	0.697939	0.39814	0.712897	[29663 16370] [11402 34480]
14	RFC	ANOVA	0.697721	0.678284	0.750338	0.750338	0.645276	0.697807	0.397788	0.712494	[29704 16329] [11455 34427]
15	GBC	ANOVA	0.698308	0.682212	0.740617	0.740617	0.656138	0.698378	0.398156	0.710216	[30204 15829] [11901 33981]
16	LRC	RELIEF	0.736213	0.72241	0.765834	0.765834	0.706689	0.736261	0.473329	0.743488	[32531 13502] [10744 35138]
17	ETC	RELIEF	0.769733	0.749098	0.810013	0.810013	0.729585	0.769799	0.541312	0.778365	[33585 12448] [8717 37165]
18	RFC	RELIEF	0.772562	0.745609	0.826294	0.826294	0.719006	0.77265	0.548411	0.783881	[33098 12935] [7970 37912]
19	GBC	RELIEF	0.745602	0.716632	0.811081	0.811081	0.680338	0.745709	0.49562	0.760937	[31318 14715] [8668 37214]

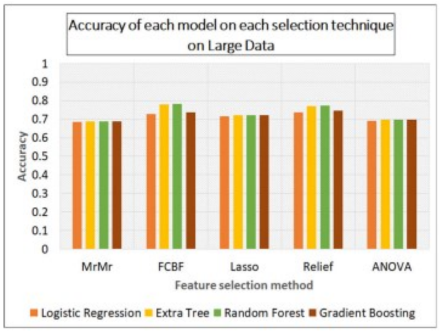


FIGURE 6. Accuracy of models on each technique for large data.

VII. CONCLUSION

The implementation of a Feature-Optimized Machine Learning Approach for cardiovascular disease (CVD) detection significantly enhances the accuracy, efficiency, and scalability of predictive healthcare systems. By integrating advanced feature selection techniques, the proposed system reduces computational complexity while improving model performance, ensuring that only the most relevant clinical parameters are used for diagnosis. Compared to traditional methods, this approach minimizes dependency on extensive medical evaluations and enables faster, automated screening for early CVD detection.

Despite these advancements, challenges such as dataset variability, real-world validation, and model interpretability require further research. Future improvements may focus on integrating explainable AI (XAI) techniques, real-time mobile health applications, and personalized predictive models to enhance clinical decision-making. With continuous advancements, machine learning-driven CVD detection has the potential to revolutionize early diagnosis, making cardiovascular risk assessment more accessible, reliable, and effective in preventing life-threatening complications.

REFERENCES

1. A. K. Gárate-Escamila, A. H. El Hassani, and E. Andrès, ‘‘Classification models for heart disease prediction using feature selection and PCA,’’ Informat. Med. Unlocked, vol. 19, Jan.

2020, Art. no. 100330, doi:10.1016/j.imu.2020.100330.

2. V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain, ‘‘An artificial intelligence model for heart disease detection using machine learning algorithms,’’ Healthcare Anal., vol. 2, Nov. 2022, Art. no. 100016, doi: 10.1016/j.health.2022.100016.
3. M. Ganesan and N. Sivakumar, ‘‘IoT based heart disease prediction and diagnosis model for healthcare using machine learning models,’’ in Proc. IEEE Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN), Mar. 2019, pp. 1–5, doi: 10.1109/ICSCAN.2019.8878850.
4. D. P. Isravel, S. V. P. Darcini, and S. Silas, ‘‘Improved heart disease diagnostic IoT model using machine learning techniques,’’ Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 4442–4446, 2020.
5. I. S. G. Brites, L. M. da Silva, J. L. V. Barbosa, S. J. Rigo, S. D. Correia, and V. R. Q. Leithardt, ‘‘Machine learning and IoT applied to cardiovascular diseases identification through heart sounds: A literature review,’’ Informatics, vol. 8, no. 4, p. 73, Oct. 2021, doi: 10.3390/informatics8040073.
6. D. T. Thai, Q. T. Minh, and P. H. Phung, ‘‘Toward an IoT-based expert system for heart disease diagnosis,’’ in Proc. 28th Mod. Artif. Intell. Cogn. Sci. Conf. (MAICS), 2017, pp. 157–164.
7. B. Padmaja, C. Srinidhi, K. Sindhu, K. Vanaja, N. M. Deepika, and E. K. R. Patro, ‘‘Early and accurate prediction of heart disease using machine learning model,’’ Turkish J. Comput. Math. Educ., vol. 12, no. 6,pp. 4516–4528, 2021.
8. S. Anitha and N. Sridevi, Heart Disease Prediction Using Data Mining Techniques S Anitha, N Sridevi to Cite This Version, document HAL Id Hal02196156, 2019. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02196156/document
9. R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, ‘‘Prediction of heart disease using a combination of machine learning and deep learning,’’ Comput. Intell. Neurosci., vol. 2021, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.
10. H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, ‘‘Heart disease prediction using machine learning algorithms,’’ IOP Conf., Mater. Sci. Eng., vol. 1022, no. 1, Jan. 2021, Art. no. 012072, doi: 10.1088/1757-899x/1022/1/012072.
11. B. Pavithra and V. Rajalakshmi, ‘‘Heart disease detection using machine learning algorithms,’’ in Proc. Int. Conf. Emerg. Current Trends Comput. Expert Technol., vol. 35, 2020, pp. 1131–1137, doi: 10.1007/978-3-030- 32150-5_115.
12. N. Louridi, S. Douzi, and B. El Ouahidi, ‘‘Machine learning-based identification of patients with a cardiovascular defect,’’ J. Big Data, vol. 8, no. 1, pp. 1–5, Dec. 2021, doi: 10.1186/s40537-021-00524-9.
13. P. Singh, G. K. Pal, and S. Gangwar, ‘‘Prediction of cardiovascular disease using feature selection techniques,’’ Int. J. Comput. Theory Eng., vol. 14, no. 3, pp. 97–103, 2022, doi: 10.7763/ijcte.2022.v14.1316.
14. M. Swathy and K. Saruladha, ‘‘A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques,’’ ICT Exp., vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.
15. D. Vaddella, C. Sruthi, B. K. Chowdary, and S.-R. Subbareddy, ‘‘Prediction of heart disease using machine learning techniques,’’ Restaur. Bus., vol. 118, no. 1, pp. 125–129, 2019, doi: 10.26643/rb.v118i1.7621.