Optimized Risk Assessment Model for Predicting Cardiac Disorders Using AI

R Vennela[1], G Viswanath[2], G Swapna[3], K Bhaskar [4]

[1] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: rallapadivennelar@gmail.com , ORCID-ID: 0009-0008-8939-3504

[2] Associate Professor, Dept. of AI & ML, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: viswag111@gmail.com , ORCID-ID: 0009-0001-7822-4739

[3] Assistant Professor, Apollo institute of pharmaceutical sciences, The Apollo University, Chittoor, India Email: swapnagv111@gmail.com , ORCID-ID: 0000-0002-9340-4148

[4] Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: bhaskark.mca@gmail.com , ORCID-ID: 0009-0000-3309-4240

Area/Section: Engineering with Medical Background
Type of the Paper: Regular Paper
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed in: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.15493550
Google Scholar Citation: IJHSP

How to Cite this Paper:

Vennela, R., Viswanath, G., Swapna, G. & Bhaskar, K. (2025). Optimized Risk Assessment Model for Predicting Cardiac Disorders Using AI. *International Journal of Health Sciences and Pharmacy (IJHSP)*, *9*(1), 140-159. DOI: https://doi.org/10.5281/zenodo.15493550

International Journal of Health Sciences and Pharmacy (IJHSP)
A Refereed International Journal of Srinivas University, India.

Crossref DOI: https://doi.org/10.47992/IJHSP.2581.6411.0138

Received on: 16/04/2025
Published on: 23/05/2025

Optimized Risk Assessment Model for Predicting Cardiac Disorders Using AI

R Vennela[1], G Viswanath[2], G Swapna[3], K Bhaskar [4]

[1] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: rallapadivennelar@gmail.com , ORCID-ID: 0009-0008-8939-3504

[2] Associate Professor, Dept. of AI & ML, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: viswag111@gmail.com , ORCID-ID: 0009-0001-7822-4739

[3] Assistant Professor, Apollo institute of pharmaceutical sciences, The Apollo University, Chittoor, India Email: swapnagv111@gmail.com , ORCID-ID: 0000-0002-9340-4148

[4] Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: bhaskark.mca@gmail.com , ORCID-ID: 0009-0000-3309-4240

ABSTRACT

Cardiovascular disease (CVD) remains the leading cause of mortality worldwide, with the World Health Organization reporting over 19.1 million deaths attributed to CVD in 2022—accounting for approximately 33% of global fatalities. Early diagnosis and effective prediction play a crucial role in reducing such numbers, especially through the integration of artificial intelligence (AI) techniques. Electrocardiogram (ECG) data, widely used in clinical settings, offers valuable insight for CVD detection, yet the challenge lies in selecting the most relevant features from this complex data. Addressing this challenge, various feature selection (FS) techniques have been implemented to optimize classification accuracy. These techniques include Analysis of Variance (ANOVA) FS with Particle Swarm Optimization (PSO), Minimum Redundancy Maximum Relevance (MRMR) FS with PSO, Least Absolute Shrinkage and Selection Operator (LASSO) FS with PSO, Feature Correlation Based Technique (FCBT) FS with PSO, and ReliefF FS with PSO. These FS strategies were combined with machine learning algorithms to identify meaningful cardiac-related patterns and remove redundant or noisy features. The models were trained and validated using benchmark datasets such as the Heart Health Data Database (HHDD) and Behavioral Risk Factor Surveillance System (BRFSS), ensuring diverse data representation for improved generalizability. Among the classifiers evaluated, a Voting Classifier—an ensemble learning method that combines multiple base models—exhibited the highest predictive accuracy, reliability, and robustness across all feature selection methods and datasets. This approach leverages the strengths of individual classifiers to provide a more stable and accurate prediction of cardiac disorders. The outcomes clearly emphasize the effectiveness of combining optimized feature selection strategies with ensemble-based classification in enhancing AI-powered diagnosis. This fusion of statistical selection and advanced machine learning enables more precise and scalable solutions for cardiac risk prediction, contributing significantly to improved clinical decision support systems and better patient outcomes.

Keywords: Cardiovascular disease, feature selection, optimization, machine learning, Voting classifier

1. INTRODUCTION:

Public health continues to be a essential worldwide task. the "world health organization (WHO)" asserts that health constitutes a essential human right. though, the arena persists in confronting the peril of numerous epidemic diseases, resulting in vast fatality quotes. "chronic disease (CDs)" notably make contributions to worldwide mortality, affecting a widespread segment of the population. those ailments, encompassing cancer, diabetes, stroke, Parkinson's disorder, and "cardiovascular disease (CVD)", are incurable and exhibit prolonged endurance within the frame relative to different problems. bad nutritional practices, tobacco use, excessive alcohol intake, and sedentary behaviors significantly make contributions to the increase of chronic diseases. inside the usa, around fifty percent of the populace endures at the least one persistent disease, and greater than 80 percentage come across monetary pressure associated with healthcare charges [1].

greater lifestyle picks markedly lessen the prevalence of continual diseases. the united states incurs a enormous fee, dedicating nearly $2.70 trillion each yr to the management of continual diseases, or 18% of its gross domestic product. "Cardiovascular disease (CVD)" is the major purpose of mortality, accounting for round 647,000 deaths annually. Likewise, several countries come across big problems related to "cardiovascular disease (CVD)". chronic diseases constitute 86.5% of fatalities in China, highlighting their significant have an impact on [3].

Cardiovascular illnesses have grow to be a leading reason of demise global, ensuing in kind of 19.1 million fatalities in 2022 and constituting 33% of all deaths, as reported through the WHO. "Cardiovascular ailment (CVD)" outcomes in 200,000 fatalities each year in Pakistan, with mortality costs at the upward thrust. the "european Society of Cardiology (ESC)" estimates that 26.five million adults in Europe are tormented by "cardiovascular disease (CVD)", with 3.8 million new cases diagnosed annually. Alarmingly, 50–55% of cardiovascular sickness sufferers do now not continue to exist past 365 days, putting big strain on healthcare structures. furthermore, around four% of global healthcare budget are special for cardiovascular disease remedy [4], [5].

Cardiovascular disease symptoms encompass physical weak point, dyspnea, edema within the legs, and tiredness [6]. danger elements like hypertension, hyperlipidemia, tobacco use, weight problems, and insufficient bodily pastime contribute to the ailment's occurrence. Cardiovascular sickness consists of congenital heart defects, congestive heart failure, and arrhythmias. conventional diagnostic tactics for "cardiovascular disease (CVD)" were tricky; but, improvements in machine learning and medical data mining have facilitated early identity and danger evaluation, presenting a promising technique to relieve the global burden of this lifestyles-threatening condition [7], [8].

2. OBJECTIVES:

This model aims to improve the early prediction of cardiac disorders using intelligent data-driven approaches, ensuring higher accuracy, better feature optimization, and enhanced clinical relevance in detection systems.

(1) To implement and evaluate multiple feature selection techniques including ANOVA-PSO, MRMR-PSO, LASSO-PSO, FCBT-PSO, and ReliefF-PSO to identify the most relevant and non-redundant ECG features for cardiac risk prediction.

(2) To utilize diverse health datasets like HHDD and BRFSS to train and validate the model, ensuring robustness, generalizability, and applicability across varied patient populations and real-world health data sources.

(3) To develop an ensemble-based classification system using a Voting Classifier, combining multiple machine learning models to enhance prediction accuracy, precision, and reliability in identifying cardiovascular disorders.

3. REVIEW OF LITERATURE/ RELATED WORKS:

Recent breakthroughs in machine learning and deep learning methodologies have markedly enhanced the prediction and prognosis of "cardiovascular diseases (CVD)", overcoming the constraints of conventional techniques. Bharti et al. [9] investigated the amalgamation of device learning and deep learning models, accomplishing improved accuracy in cardiovascular disorder prediction by the synthesis of various algorithms. Jindal et al. [32] examined the utilization of various machine learning techniques, together with "support Vector Machines (SVM), decision trees, and Random Forests", for predicting cardiac disease, emphasizing their efficacy in coping with extensive datasets and enhancing diagnostic accuracy. Pavithra and Rajalakshmi [11] highlighted the efficacy of ensemble learning techniques for cardiovascular disease detection, showcasing better performance through algorithmic collaboration. Louridi et al. [34] concentrated on identifying cardiovascular troubles through system learning models, highlighting its capability to investigate complicated healthcare datasets and derive good sized styles for patient analysis.

Singh et al. [13] investigated feature selection methodologies to enhance the precision of cardiovascular ailment prediction fashions. Their studies proven that strategies like "Recursive feature elimination (RFE) and principal component analysis (PCA)" markedly superior classification efficacy through diminishing facts dimensionality. Swathy and Saruladha [36] carried out a comparative analysis of machine learning and deep learning methodologies for cardiovascular disorder prediction, concluding that deep learning models, including "Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)", exhibited advanced accuracy and robustness whilst processing massive clinical datasets. Vaddella et al. [15] investigated multiple machine learning methodologies, inclusive of "k-Nearest neighbors (KNN) and Naive Bayes", for predicting cardiac infection, emphasizing their relevance in actual-time scientific environments as a result of their simplicity and computational efficiency.

Ramalingam et al. [22] performed an extensive survey of machine learning packages in heart disorder prediction, highlighting the significance of predictive models in reducing demise fees. Their findings emphasized the importance of combining area-particular information with sophisticated computational models to overcome the shortcomings of traditional diagnostic techniques. that research together show the transformational capacity of system learning and deep learning in healthcare, especially regarding cardiovascular disorder analysis and threat assessment. They emphasize the importance of feature selection, model optimization, and algorithmic integration to beautify predictive accuracy and provide early intervention methods for stepped forward patient results.

Table 1: Literature Survey Comparison Table

Sl. No	Area & Focus of the Research	The result of the Research	Reference
1	Combining machine and deep learning models for CVD prediction	Achieved improved accuracy by integrating diverse AI-based algorithms	R. Bharti, A. Khamparia, M. Shabaz et. al.., (2021). [9]
2	Explored SVM, decision trees, and random forests in CVD	Handled large datasets with increased diagnostic prediction accuracy	H. Jindal, S. Agrawal, R. Khera et. al., (2021) [32]
3	Focused on ensemble learning for cardiovascular disease detection	Enhanced prediction performance using algorithm collaboration techniques	B. Pavithra and V. Rajalakshmi (2020) [11]
4	Studied feature selection to improve CVD model accuracy	Reduced dimensionality and improved classification effectiveness significantly	P. Singh, G. K. Pal, and S. Gangwar (2022) [13]
5	Compared deep and machine learning models on CVD data	CNNs and RNNs outperformed traditional models in accuracy	M. Swathy and K. Saruladha. (2022) [36]

4. MATERIALS AND METHODS:

The counseled system seeks to enhance the prediction of "cardiovascular diseases (CVD)" via incorporating sophisticated characteristic selection techniques along various machine learning algorithms. The device employs characteristic selection strategies like "ANOVA FS, MRMR FS, Lasso FS, FCBT (Correlation) FS, and ReliefF FS", all greater by using Particle Swarm Optimization (PSO) [21] to ascertain the most pertinent features from the dataset. The characteristics are further processed with the aid of diverse category methods, including "Logistic Regression [16], Random forest [17], ExtraTree [18], and Gradient Boosting [19]", to forecast CVD chance. A vote casting Classifier that integrates "AdaBoost with DecisionTree and ExtraTree" classifiers is applied to decorate prediction accuracy. This ensemble technique ensures strong performance by consolidating the predictions of diverse models, subsequently enhancing the system's reliability and its ability to manipulate difficult healthcare records. The device seeks to beautify the accuracy and interpretability of predictions for cardiovascular ailment threat evaluation by incorporating feature selection and different machine learning methodologies.

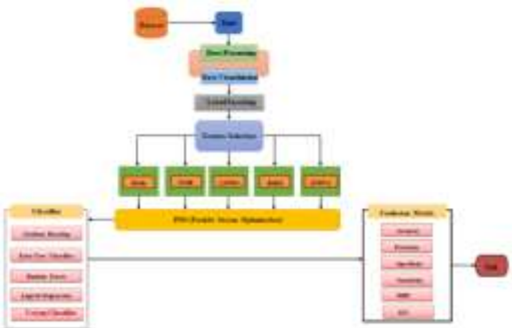


Fig 1: Proposed Architecture

The system architecture (fig. 1) comprises a sequence of strategies for the detection of cardiovascular ailment. The system begins with records preprocessing, visualization, and label encoding. in the end, feature selection is performed using various methodologies. the chosen capabilities are entered into numerous classifiers, and their outputs are aggregated through a vote casting classifier. ultimately, the device assesses achievement through various measures.

4.1 Dataset Collection:

This work employs two datasets: the BRFSS dataset, comprising 2000 entries and 22 features associated with health behaviors and cardiovascular hazard factors, and the HHDD dataset, together with 1190 entries and 12 functions targeted on medical characteristics for heart disease prediction. both databases provide thorough evaluation for cardiovascular threat assessment.

BRFSS

This look at utilizes the "Behavioral risk factor Surveillance system (BRFSS)" dataset, comprising 2000 entries and 22 attributes. The traits encompass "HeartDiseaseorAttack, HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, Diabetes, PhysActivity, culmination, veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCost, GenHlth, MentHlth, PhysHlth, DiffWalk, sex, Age, training, and income". the collection offers insights into many health behaviors and situations, facilitating the prediction and examine of cardiovascular disease risks.

Table 2: Dataset Collection Table – BRFSS

	HeartDiseaseorAttack	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	Diabetes
0	0.0	1.0	1.0	1.0	40.0	1.0	0.0	0.0
1	0.0	0.0	0.0	0.0	25.0	1.0	0.0	0.0
2	0.0	1.0	1.0	1.0	28.0	0.0	0.0	0.0
3	0.0	1.0	0.0	1.0	27.0	0.0	0.0	0.0
4	0.0	1.0	1.0	1.0	24.0	0.0	0.0	0.0

HHDD

This observe utilizes the “heart disease diagnosis (HHDD) dataset”, which contains 1190 entries and 12 characteristics. The capabilities encompass “age, sex, sort of chest ache, resting blood pressure, levels of cholesterol, fasting blood sugar, resting electrocardiogram, most heart price, exercise-induced angina, oldpeak, ST segment slope, and goal variable”. This dataset is supposed for forecasting the prevalence of heart disorder primarily based on various health-related indicators, facilitating the exam of cardiovascular chance elements and the introduction of heart disorder prediction models.

Table 3: Dataset Collection Table – HHDD

	age	sex	chest pain type	resting bp s	cholesterol	fasting blood sugar	resting ecg	max heart rate	exercise angina	oldpeak	ST slope
0	40	1	2	140	289	0	0	172	0	0.0	1
1	49	0	3	160	180	0	0	156	0	1.0	2
2	37	1	2	130	283	0	1	98	0	0.0	1
3	48	0	4	138	214	0	0	108	1	1.5	2
4	54	1	3	150	195	0	0	122	0	0.0	1

4.2 Pre-Processing:

During the pre-processing segment, we focus on preparing the dataset for modeling. This encompasses statistics cleansing, showing great institutions, encoding specific labels, and executing characteristic choice to guarantee enter for the predictive version.

4.2.1 Data Processing

This stage ensures that the datasets are sanitized and pertinent for analysis. data cleaning encompasses handling absent values, identifying and rectifying discrepancies, and mitigating outliers to improve data integrity. duplicate entries are eliminated to save you bias. inappropriate columns are eliminated to beautify the prediction undertaking, thereby decreasing dimensionality and computing complexity. as an instance, functions showing no variability or negligible connection with the target variable are overlooked, thereby enhancing the dataset for next analysis.

4.2.2 Data Visualization

data visualization facilitates comprehension of the dataset's structure and the interrelationships amongst capabilities. A sample final results visualization, consisting of bar charts or pie charts, illustrates the distribution of the goal variable, offering insights on facts imbalance. The correlation matrix delineates links among attributes, emphasizing couples with sturdy correlations. functions exhibiting extensive multicollinearity may be identified for elimination or change. those visual insights decorate the dataset, making certain the selection of pertinent and massive features.

4.2.3 Label Encoding

Label encoding is applied to transform class variables into numerical values, therefore permitting their utility in machine learning techniques. every distinct category internal a feature is allocated a numeric code, making sure interoperability with fashions necessitating numerical input. for example, binary variables like "sex" are assigned the values 0 and 1. This stage keeps the integrity of express facts even as enabling algorithms to procedure it successfully, making sure that the functions are appropriately represented within the modeling procedure.

4.2.3 Feature Selection

Feature selection determines the maximum pertinent characteristics for the predictive process, improving version efficacy and mitigating overfitting. Techniques like “correlation evaluation, ANOVA, FCBT, LASSO, and ReliefF with PSO” [21] are applied to prioritize and preserve vital features. Beside the point or duplicated residences are removed. This procedure improves the computational performance and predictive capability of machine learning models by ensuring the dataset has just the most vast capabilities for unique predictions.

4.3 Training & Testing

The training and testing manner includes dividing the preprocessed dataset into two subsets: one precise for version training and the opposite for assessment. The education information is applied to construct the predictive model by means of discerning patterns and correlations among traits and the goal variable. The unseen testing information is employed to evaluate the version's overall performance and generalization capability. This guarantees the version's capacity to precisely forecast consequences on novel, unobserved facts, subsequently augmenting its dependability.

4.4 Algorithms:

LR: Logistic Regression is hired to correctly version the affiliation among capabilities and the target variable. features are selected utilising “ANOVA with PSO, MRMR with PSO, Lasso with PSO, FCBT (Correlation) with PSO, and ReliefF with PSO to check the most pertinent predictors”. This technique ensures the incorporation of optimum feature subsets, facilitating a honestly described linear decision boundary for binary type [16].

RF: Random Forest, an ensemble technique comprising decision trees, is employed to manage excessive-dimensional datasets with stability and resilience. feature selection methodologies, inclusive of “ANOVA with PSO, MRMR with PSO, Lasso with PSO, FCBT (Correlation) with PSO, and ReliefF with PSO”, beautify input functions to enhance accuracy. This complements the version's ability to parent complex interactions among chosen functions while lowering overfitting [17].

ExtraTree: ExtraTreeis utilized for its computational efficiency and ability to generate randomized decision trees. utilising “ANOVA with PSO, MRMR with PSO, Lasso with PSO, FCBT (Correlation) with PSO, and ReliefF with PSO” for characteristic selection, it efficiently strategies subtle inputs to model non-linear relationships. The randomization contributes to variance reduction throughout the evaluation of important functions [18].

GB: Gradient Boosting sequentially reduces mistakes to decorate predictions. Optimized characteristic subsets produced with “ANOVA with PSO, MRMR with PSO, Lasso with PSO, FCBT (Correlation) with PSO, and ReliefF with PSO” improve performance. This iterative approach ensures the collection of nuanced information styles, enhancing accuracy and resilience across selected capabilities [19].

VC: The Voting Classifier integrates AdaBoost DecisionTree and ExtraTree to form a resilient ensemble. characteristic selection techniques including “ANOVA with PSO, MRMR with PSO, Lasso with PSO, FCBT (Correlation) with PSO, and ReliefF with PSO yield greater and optimized inputs”. by consolidating predictions, it achieves advanced accuracy, harmonizing the benefits of its exceptional classifiers for optimum effects.

5. RESULTS AND DISCUSSION:

Accuracy: The accuracy of the test concerns its ability to correctly distinguish between patient and healthy cases. In order to assess the accuracy of the test, one must calculate the ratio of real positives and real negatives in all evaluated cases. This can be mathematically expressed as:

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

Precision: The accuracy evaluates the share of precisely classified cases among cases identified as positive. As a result, the formula for calculating accuracy is expressed:

"$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$ "

Recall: The invocation is a meter in machine learning that evaluates the ability of the model to recognize all relevant cases of a particular class. It is the proportion of precisely predicted positive observations of total real positives and offers insight into the efficiency of the model in identifying the occurrence of a particular class.

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

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

"$F1\,Score = 2 * \frac{Recall\ X\ Precision}{Recall + Precision} * 100(1)$"

AUC-ROC Curve: The AUC-ROC Curve is a metric for comparing class overall performance throughout different threshold levels. ROC plots the true positive rate as opposed to the false positive rate. The AUC measures the model's ordinary ability to differentiate across lessons, with a higher AUC signifying advanced version overall performance.

"$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_{i+1} + TPR_i}{2} (4)$"

Sensitivity: Sensitivity quantifies the efficacy of a check or equipment in detecting a condition in an man or woman. it is decided by contrasting the wide variety of individuals who take a look at fantastic for a circumstance with the real occurrence of that disease.

"$Specificity = \frac{TP}{(TP+FN)} (5)$"

Specificity: The calculation involves determining the range of people who test poor for a situation and dividing it via the total variety of individuals with out the ailment, encompassing both those who tested negative and the false positives—individuals who examined wonderful but do not have the disease.

"$Specificity = \frac{TN}{(TN + FP)} (6)$"

MCC: The Matthews coefficient, or Matthews correlation coefficient (MCC), is a performance indicator utilized for binary classifiers in machine learning. It assesses the correlation among anticipated and actual binary consequences by using comparing all 4 additives of a confusion matrix.

"$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN}} (7)$"

Tables (1 to ten) assess the performance metrics—F1-score, precision, AUC, accuracy, MCC, specificity, and sensitivity—for every approach. The voting Classifier often surpasses all other algorithms across all characteristic selection strategies. The tables provide a comparative exam of the metrics for the opportunity strategies.

Table 4: Performance Evaluation Metrics for Anova FS with PSO – BRFSS

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Anova-PSO LR	0.864	0.961	0.907	0.710	0.200	0.864	0.140
Anova-PSO RF	0.846	0.925	0.882	0.857	0.198	0.846	0.074
Anova-PSO ET	0.850	0.915	0.878	0.872	0.244	0.850	0.141
Anova-PSO GB	0.868	0.959	0.907	0.732	0.224	0.868	0.186
Anova-PSO Voting Classifier	**0.918**	**0.950**	**0.929**	**0.621**	**0.529**	**0.918**	**0.588**

Graph 1: Comparison Graphs for Anova FS with PSO – BRFSS

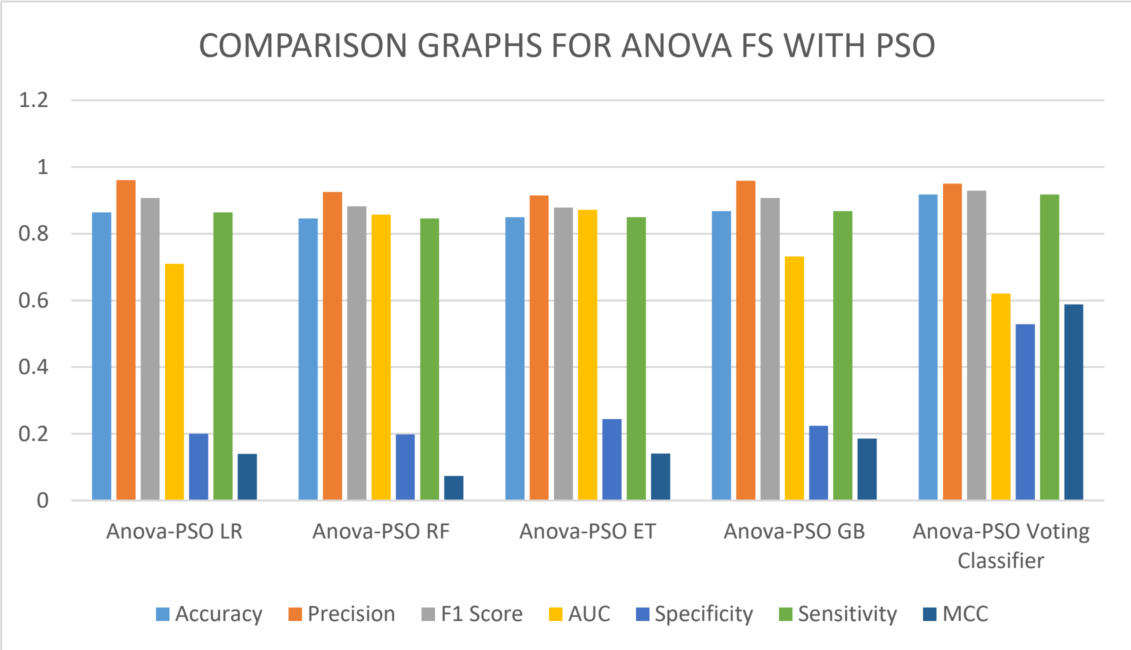


Table 5: Performance Evaluation Metrics for MRMR FS with PSO – BRFSS

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
MRMR-PSO LR	0.868	0.964	0.911	0.770	0.173	0.868	0.093
MRMR-PSO RF	0.879	0.948	0.907	0.881	0.272	0.879	0.257
MRMR-PSO ET	0.875	0.941	0.903	0.891	0.272	0.875	0.240
MRMR-PSO GB	0.875	0.959	0.911	0.786	0.223	0.875	0.194
MRMR-PSO Voting Classifier	**0.932**	**0.963**	**0.942**	**0.687**	**0.549**	**0.932**	**0.649**

Graph 2: Comparison Graphs for MRMR FS with PSO – BRFSS

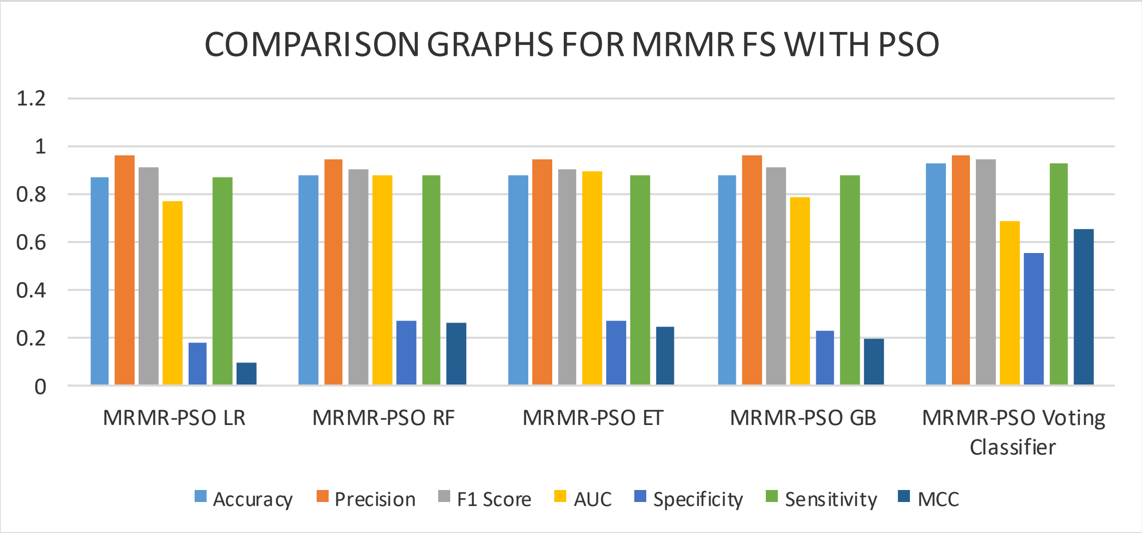


Table 6: Performance Evaluation Metrics for Lasso FS with PSO – BRFSS

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Lasso-PSO LR	0.868	0.964	0.911	0.770	0.173	0.868	0.093
Lasso-PSO RF	0.879	0.948	0.907	0.881	0.272	0.879	0.257
Lasso-PSO ET	0.871	0.943	0.902	0.891	0.247	0.871	0.202
Lasso-PSO GB	0.875	0.959	0.911	0.786	0.223	0.875	0.194
Lasso-PSO Voting Classifier	**0.932**	**0.963**	**0.942**	**0.686**	**0.549**	**0.932**	**0.649**

Graph 3: Comparison Graphs for Lasso FS with PSO – BRFSS

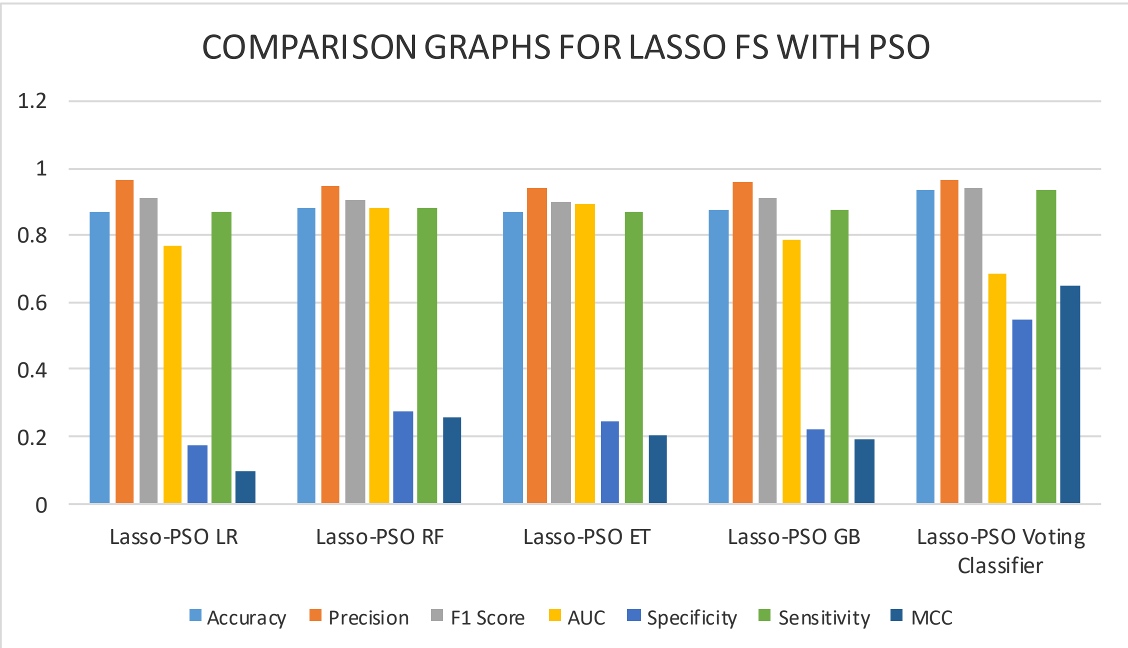


Table 7: Performance Evaluation Metrics for FCBT FS with PSO – BRFSS

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
FCBT-PSO LR	0.847	0.985	0.906	0.758	0.208	0.847	0.182
FCBT-PSO RF	0.841	0.962	0.893	0.780	0.223	0.841	0.149
FCBT-PSO ET	0.841	0.971	0.897	0.781	0.207	0.841	0.129
FCBT-PSO GB	0.847	0.975	0.901	0.762	0.224	0.847	0.192
FCBT-PSO Voting Classifier	**0.841**	**0.962**	**0.893**	**0.780**	**0.223**	**0.841**	**0.149**

Graph 4: Comparison Graphs for FCBT FS with PSO – BRFSS

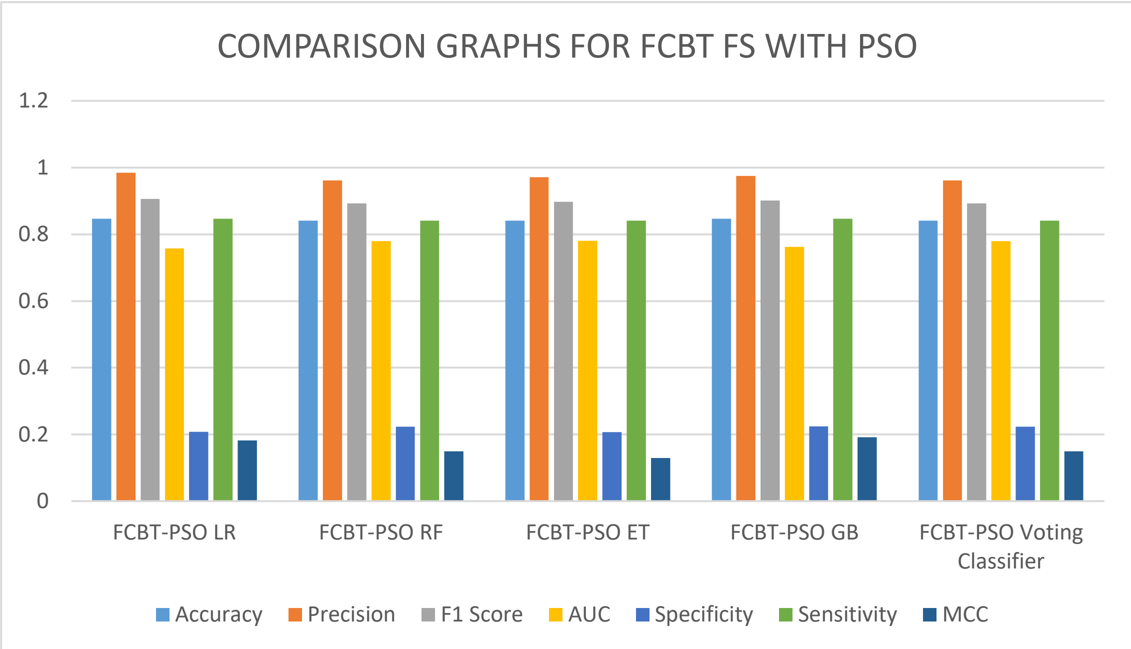


Table 8: Performance Evaluation Metrics for Relief FS with PSO – BRFSS

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Relief-PSO LR	0.832	0.833	0.832	0.862	0.833	0.832	0.665
Relief-PSO RF	0.844	0.845	0.844	0.891	0.845	0.844	0.690
Relief-PSO ET	0.844	0.845	0.844	0.892	0.845	0.844	0.690
Relief-PSO GB	0.850	0.850	0.850	0.880	0.850	0.850	0.700
Relief-PSO Voting Classifier	**0.856**	**0.856**	**0.856**	**0.866**	**0.857**	**0.856**	**0.713**

Graph 5: Comparison Graphs for Relief FS with PSO – BRFSS

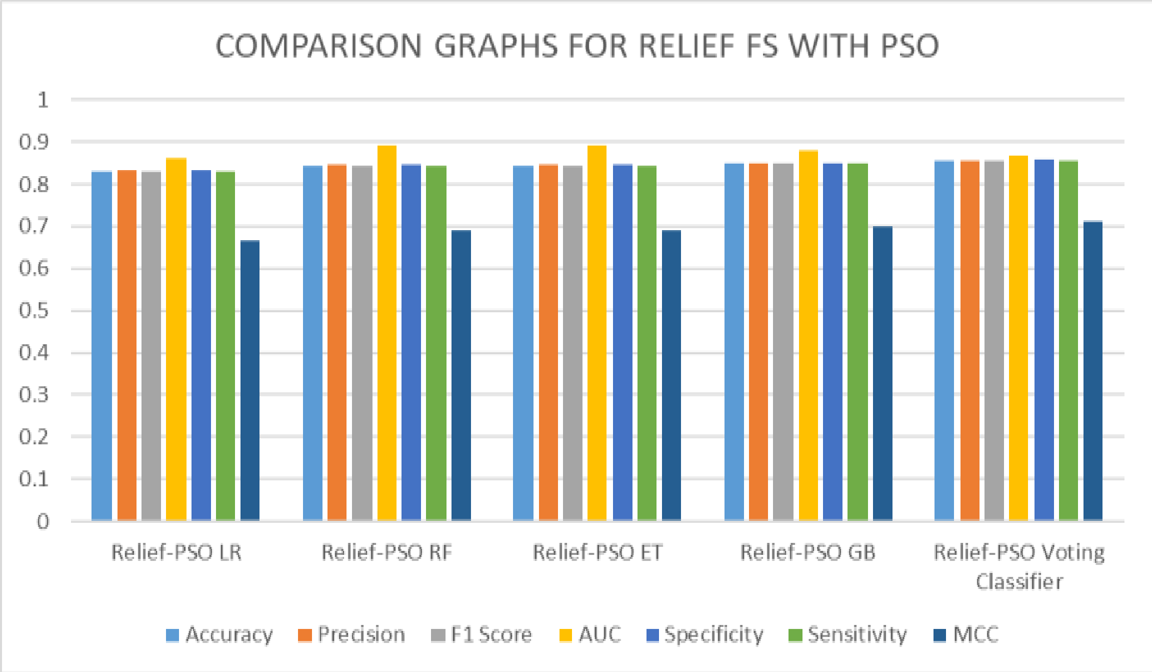


Table 9: Performance Evaluation Metrics for Anova FS with PSO – HHDD

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Anova-PSO LR	0.838	0.840	0.838	0.883	0.839	0.838	0.678
Anova-PSO RF	0.832	0.832	0.832	0.973	0.832	0.832	0.665
Anova-PSO ET	0.814	0.814	0.814	0.978	0.814	0.814	0.629
Anova-PSO GB	0.832	0.846	0.833	0.921	0.835	0.832	0.675
Anova-PSO Voting Classifier	**0.940**	**0.941**	**0.940**	**0.873**	**0.939**	**0.940**	**0.881**

Graph 6: Comparison Graphs for Anova FS with PSO – HHDD

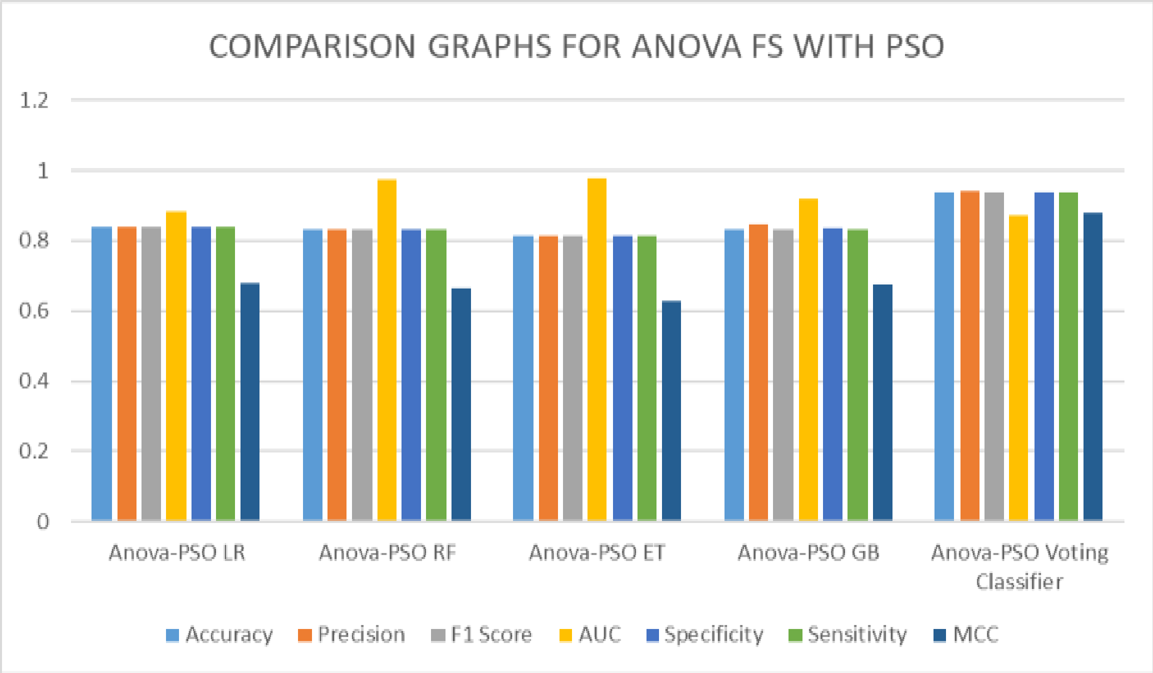


Table 10: Performance Evaluation Metrics for MRMR FS with PSO – HHDD

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
MRMR-PSO LR	0.85	0.851	0.85	0.783	0.851	0.85	0.7
MRMR-PSO RF	0.85	0.851	0.85	0.805	0.851	0.85	0.7
MRMR-PSO ET	0.85	0.851	0.85	0.805	0.851	0.85	0.7
MRMR-PSO GB	0.85	0.851	0.85	0.801	0.851	0.85	0.7
MRMR-PSO Voting Classifier	**0.85**	**0.851**	**0.85**	**0.805**	**0.851**	**0.85**	**0.7**

Graph 7: Comparison Graphs for MRMR FS with PSO – HHDD

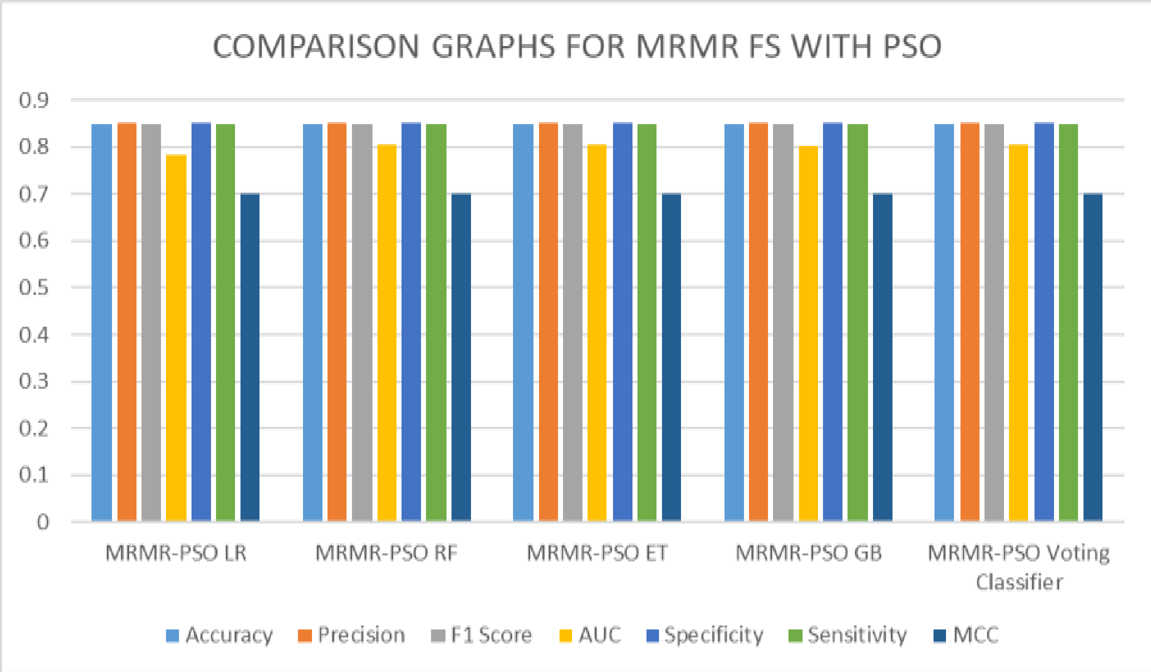


Table 11: Performance Evaluation Metrics for Lasso FS with PSO – HHDD

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Lasso-PSO LR	0.85	0.851	0.85	0.783	0.851	0.85	0.7
Lasso-PSO RF	0.85	0.851	0.85	0.805	0.851	0.85	0.7
Lasso-PSO ET	0.85	0.851	0.85	0.805	0.851	0.85	0.7
Lasso-PSO GB	0.85	0.851	0.85	0.801	0.851	0.85	0.7
Lasso-PSO Voting Classifier	**0.85**	**0.851**	**0.85**	**0.805**	**0.851**	**0.85**	**0.7**

Graph 8: Comparison Graphs for Lasso FS with PSO – HHDD

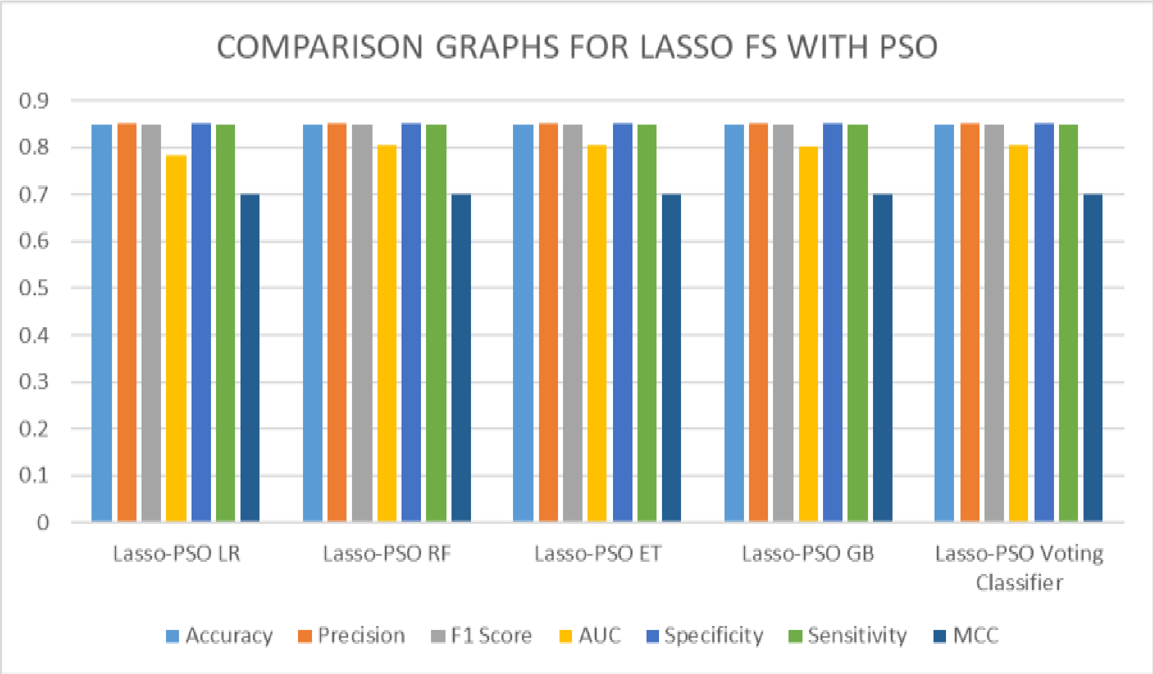


Table 12: Performance Evaluation Metrics for FCBT FS with PSO – HHDD

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
FCBT-PSO LR	0.843	0.857	0.844	0.877	0.844	0.843	0.696
FCBT-PSO RF	0.843	0.848	0.843	0.968	0.843	0.843	0.690
FCBT-PSO ET	0.859	0.859	0.859	0.974	0.859	0.859	0.718
FCBT-PSO GB	0.853	0.859	0.854	0.914	0.854	0.853	0.711
FCBT-PSO Voting Classifier	**0.948**	**0.950**	**0.948**	**0.884**	**0.948**	**0.948**	**0.897**

Graph 9: Comparison Graphs for FCBT FS with PSO – HHDD

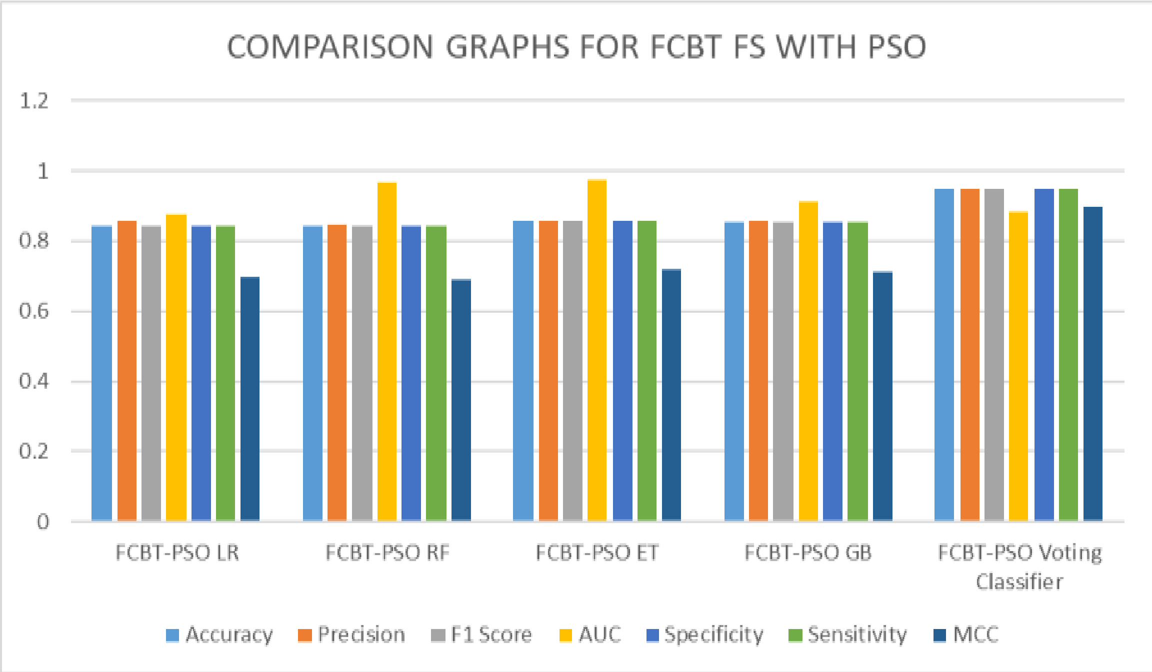
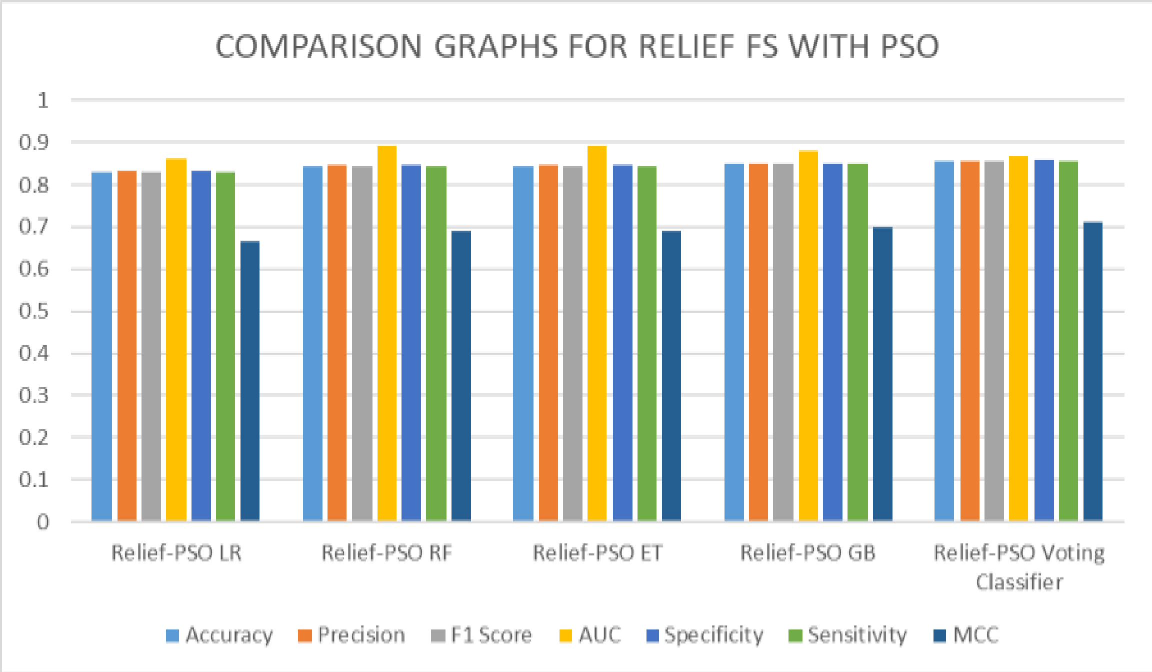


Table 13: Performance Evaluation Metrics for Relief FS with PSO – HHDD

ML Model	Accuracy	Precision	F1 Score	AUC	Specificity	Sensitivity	MCC
Relief-PSO LR	0.832	0.833	0.832	0.862	0.833	0.832	0.665
Relief-PSO RF	0.844	0.845	0.844	0.891	0.845	0.844	0.690
Relief-PSO ET	0.844	0.845	0.844	0.892	0.845	0.844	0.690
Relief-PSO GB	0.850	0.850	0.850	0.880	0.850	0.850	0.700
Relief-PSO Voting Classifier	**0.856**	**0.856**	**0.856**	**0.866**	**0.857**	**0.856**	**0.713**

Graph 10: Comparison Graphs for Relief FS with PSO – HHDD



In Graphs (1 to 10), accuracy is depicted in mild blue, precision in orange, F1 score in gray, AUC in light yellow, specificity in blue, sensitivity in green, and MCC in dark blue. compared to the alternative models, the voting Classifier has more overall performance across all feature selection techniques, achieving the very best values. The graphs above visually represent these findings.

6. CONCLUSION:

In conclusion, our thorough exam of numerous feature selection strategies and machine learning algorithms for “cardiovascular disease (CVD)” identification, utilizing the “BRFSS and HHDD” datasets [20], has underscored the efficacy of precise methodologies in reaching expanded diagnostic precision. Of the methods evaluated, the ‘voting Classifier (AdaBoost with decision Tree + ExtraTree)” had the advanced performance. This ensemble technique adeptly integrated the blessings of AdaBoost with decision tree classifiers, allowing the model to excel in classifying cardiovascular disorder danger. thru the software of state-of-the-art function selection methodologies, which includes “Anova, MRMR, Lasso, FCBT (Correlation), and ReliefF with PSO”, we efficiently optimized the input functions, as a result augmenting the model's predictive efficacy. The voting Classifier's resilient ensemble framework, which consolidates predictions from both AdaBoost selection Tree and ExtraTree, efficiently decreased bias and variance, enhancing universal overall performance. the combination of best functions and sturdy algorithmic assist yielded a version proficient in reliably figuring out cardiovascular ailment, underscoring the significance of ensemble techniques and characteristic choice techniques in enhancing category obligations. This research illustrates the efficacy of integrating data from many resources with sophisticated machine learning strategies for enhanced health risk forecasting.

Future research may look at the utilization of deep learning fashions to explain tricky characteristic interactions and enhance precision. Exploring alternate characteristic selection strategies, together with evolutionary algorithms and mutual information-primarily based techniques, may also decorate feature relevance. Integrating supplementary records sources, consisting of environmental variables, could beautify model accuracy. Superior information balancing processes, including SMOTE, alongside ensemble learning techniques like bagging or boosting, could further decorate predictive accuracy. those upgrades seek to enhance the precision and efficacy of heart disease threat prediction methods.

REFERENCES:

[1] Wu, Y., Liu, B., Wu, W., Lin, Y., Yang, C., & Wang, M. (2018). Grading glioma by radiomics with feature selection based on mutual information. *J. Ambient Intell. Humanized Comput*., 9(5), 1671–1682.

[2] Viswanath, G. (2021). Hybrid encryption framework for securing big data storage in multi-cloud environment. *Evolutionary intelligence,* 14(2), 691-698.

[3] Gupta, R., Flanagan, S., Li, C. C., Lee, M., Shivalingham, B., Maleki, S., Wheeler, H. R., & Buckland, M. E. (2013). Expanding the spectrum of IDH1 mutations in gliomas. *Modern Pathol*., 26(5), 619–625.

[4] Viswanath, G. (2023). A Real-Time Case Scenario Based On URL Phishing Detection Through Login URLS. *Material science and technology*, 22(9), 103-108.

[5] Cao, H., Erson-Omay, E. Z., Li, X., Günel, M., Moliterno, J., & Fulbright, R. K. (2020). A quantitative model based on clinically relevant MRI features differentiates lower grade gliomas and glioblastoma. *Eur. Radiol*., 30(6), 3073–3082.

[6] Viswanath, G., & Sunil Kumar Reddy, T. (2014). Enhancing power unbiased cooperative media access control protocol in manets . *International Journal of Engineering Inventions*, 4(9), 8-12.

[7] Padmaja, B., Srinidhi, C., Sindhu, K., Vanaja, K., Deepika, N. M., & Patro, E. K. R. (2021). Early and accurate prediction of heart disease using machine learning model. *Turkish J. Comput. Math. Educ.*, 12(6), 4516–4528.

[8] Viswanath, G. (2024). Artificial Intelligence-driven Frameworks for Fostering Active Participation and Learning in Language Classrooms. *International Journal of Interpreting Enigma Engineers (IJIEE),* 1(3), 23-32.

[9] L. Zou, L., X. Luo, X., Y. Zhang, Y., X. Yang, X., & and X. Wang, X. (2023). HC-DTTSVM: A network intrusion detection method based on decision tree twin support vector machine and hierarchical clustering. *IEEE Access*, 11(1), 21404–21416.

[10] Viswanath, G. (2024). Improved LightGBM Model Performance Analysis and Comparison For Coronary Heart Disease Prediction. *International Journal of Information Technology and Computer Engineering*, 12(3), 658-672.

[11] Pavithra, B., & Rajalakshmi, V. (2020). Heart disease detection using machine learning algorithms. *in Proc. Int. Conf. Emerg. Current Trends Comput. Expert Technol*., 35(3), 1131–1137.

[12] Viswanath, G. (2024). Hybrid Feature Extraction With Machine Learning To Identify Network Attacks. *International Journal of HRM and Organizational Behavior*, 12(3), 217-228.

[13] Singh, P., Pal, G. K., & Gangwar, S. (2022). Prediction of cardiovascular disease using feature selection techniques. *Int. J. Comput. Theory Eng*., 14(3), 97–103.

[14] Viswanath, G. (2023). A Real Time Online Food Ording Application Based Django Restfull Framework. *Juni Khyat*, 13(9), 154-162

[15] Vaddella, D., Sruthi, C., Chowdary, B. K., & Subbareddy, S. R. (2019). Predic tion of heart disease using machine learning techniques. *Restaur. Bus.*, 118(1), 125–129.

[16] Ganesh, A. G. B., Ganesh, A., Srinivas, C., & Mensinkal, K. (2022). Logis tic regression technique for prediction of cardiovascular disease. *Global Transitions Proc.*, 3(1), 127–130.

[17] Pal, M., & Parija, S. (2021). Prediction of heart diseases using random forest. *J. Phys., Conf. Ser.*, 1817(2021), 1-9.

[18] Hameed, M. M., AlOmar, M. K., Khaleel, F., & Al-Ansari, N. (2021). An extra tree regression model for discharge coefficient prediction: Novel, practical applications in the hydraulic sector and future research direc tions. *Math. Problems Eng.*, 2021(1), 1–19.

[19] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.*, 63(1), 3–42.

[20] Mohi Uddin, K. M., Ripa, R., Yeasmin, N., Biswas, N., & Dey, S. K. (2023). Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset. *Intell.-Based Med.*, 7(2), 1-15.

[21] A. H. Shahid A. H., &and M. P. Singh, M. P. (2020). A novel approach for coronary artery disease diagnosis using hybrid particle swarm optimization based emo tional neural network. *Biocybernetics Biomed. Eng.*, 40(4), 1568–1585.

[22] Ramalingam, V. V., Dandapath, A., & Karthik Raja, M. (2018). Heart disease prediction using machine learning techniques: A survey. *Int. J. Eng. Technol.*, 7(2), 684-687.

[23] Viswanath, G. (2024). Personalized Breast Cancer Prognosis through Data Mining Innovations. *Cuestiones de Fisioterapia*, 53(2), 538-548.

[24] Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Anal.*, 2(1), 1-17.

[25] Viswanath, G. (2021). Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage. *Turkish Journal of Computer and Mathematics Education*, 12(9), 545-554.

[26] Isravel, D. P., Darcini, S. V. P., & Silas, S. (2020). Improved heart disease diagnostic IoT model using machine learning techniques. *Int. J. Sci. Technol. Res.*, 9(2), 4442–4446.

[27] Viswanath, G. (2023). A Real -Time Video Based Vehicle Classification Detection And Counting System. *Industrial Engineering Journal*, 52(9), 474-480.

[28] Thai, D. T., Minh, Q. T., & Phung, P. H. (2017). Toward an IoT-based expert system for heart disease diagnosis. *in Proc. 28th Mod. Artif. Intell. Cogn. Sci. Conf. (MAICS)*, 1964(1), 157–164.

[29] Viswanath, G. (2025). Proactive Security in Multi-Cloud Environments: A Blockchain Integrated Real-Time Anomaly Detection and Mitigation Framework. *Cuestiones de Fisioterapia*, 54(2), 392-417.

[30] Anitha, S., & Sridevi, N. (2019). Heart Disease Prediction Using Data Mining Techniques . *HAL* ,13(2) ,48-55.

[31] Viswanath, G. (2024). International Journal of Information Technology and Computer Engineering. *International Journal of Interpreting Enigma Engineers (IJIEE)*, 12(3), 647-657.

[32] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conf., Mater. Sci. Eng.*, 1022(1), 1-11.

[33] Viswanath, G. (2024). Machine Learning for IoT Device Anomaly Detection Attack Classification. *International Journal of Mechanical Engineering Research and Technology*, 16(9), 66-76.

[34] Louridi, N., Douzi, S., & El Ouahidi, B. (2021). Machine learning-based identi f ication of patients with a cardiovascular defect. *J. Big Data*, 8(1), 1–5.

[35] Viswanath, G. (2024). A Hybrid Particle Swarm Optimization And C4.5 For Network Intrusion Detection And Prevention System. *International Journal of Computing This link is disabled*, 23(1), 109-115.

[36] Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deeplearning techniques. *ICT Exp*., 8(1) , 109–116.

[37] Viswanath, G. (2014). Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS. *International Journal of Engineering Inventions,* 4(2), 08-12.

****