



OPEN ACCESS

EDITED BY

Hadi Akbarzadeh Khorshidi,
The University of Melbourne, Australia

REVIEWED BY

Salvatore Andrea Pullano,
University Magna Graecia of Catanzaro, Italy
Fan He,
The University of Melbourne, Australia

*CORRESPONDENCE

G. Hannah Grace
✉ hannahgrace.g@vit.ac.in

RECEIVED 10 April 2025

ACCEPTED 04 June 2025

PUBLISHED 19 June 2025

CITATION

Ashika T and Hannah Grace G (2025)
Enhancing heart disease prediction with
stacked ensemble and MCDM-based ranking:
an optimized RST-ML approach.
Front. Digit. Health 7:1609308.
doi: 10.3389/fdgth.2025.1609308

COPYRIGHT

© 2025 Ashika and Hannah Grace. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhancing heart disease prediction with stacked ensemble and MCDM-based ranking: an optimized RST-ML approach

T. Ashika and G. Hannah Grace*

Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology Chennai, Chennai, India

Introduction: Cardiovascular disease (CVD) is a leading global cause of death, necessitating the development of accurate diagnostic models. This study presents an Optimized Rough Set Theory-Machine Learning (RST-ML) framework that integrates Multi-Criteria Decision-Making (MCDM) for effective heart disease (HD) prediction. By utilizing RST for feature selection, the framework minimizes dimensionality while retaining essential information.

Methods: The framework employs RST to select relevant features, followed by the integration of nine ML classifiers into five stacked ensemble models through correlation analysis to enhance predictive accuracy and reduce overfitting. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) ranks the models, with weights assigned using the Mean Rank Error Correction (MEREC) method. Hyperparameter tuning for the top model, Stack-4, was conducted using GridSearchCV, identifying XGBoost (XG) as the most effective classifier. To assess scalability and generalization, the framework was evaluated using additional datasets, including chronic kidney disease (CKD), obesity levels, and breast cancer. Explainable AI (XAI) techniques were also applied to clarify feature importance and decision-making processes.

Results: Stack-4 emerged as the highest-performing model, with XGBoost achieving the best predictive accuracy. The application of XAI techniques provided insights into the model's decision-making, highlighting key features influencing predictions.

Discussion: The findings demonstrate the effectiveness of the RST-ML framework in improving HD prediction accuracy. The successful application to diverse datasets indicates strong scalability and generalization potential, making the framework a robust and scalable solution for timely diagnosis across various health conditions.

KEYWORDS

rough set theory, machine learning, correlation analysis, stacking classifier, multi-criteria decision-making, gridsearchCV

1 Introduction

CVD is a leading global cause of mortality, accounting for 17.9 million deaths annually, as estimated by the World Health Organization (WHO). CVD encompasses various heart and blood vessel conditions and is strongly associated with several modifiable risk factors, including stress, poor diet, physical inactivity, smoking, and excessive alcohol consumption. Factors such as obesity, hypertension, high cholesterol,

and tobacco use further exacerbate its severity, emphasizing the need for early detection and intervention (1). Artificial Intelligence (AI) has revolutionized medical research, with ML playing a critical role in predictive modelling for clinical diseases (2).

AI and data mining techniques facilitate early CVD risk prediction by analysing large-scale behavioural and clinical data, identifying patterns associated with disease progression, and supporting timely preventive interventions. While diagnosis may involve imaging, predictive models using ML, support pre-symptomatic risk assessment, enabling timely intervention (3, 4). Traditional statistical methods, while effective under specific data distributions, often face limitations when applied to non-linear, high-dimensional clinical datasets. These challenges have led to increased adoption of ML and DL techniques to improve predictive performance in complex healthcare scenarios (5).

ML excels in handling large-scale medical data, enabling early-stage disease prediction and reducing preventable hospitalizations. Its application enhances healthcare policies, disease prevention, and medical decision-making. Many studies have explored ML techniques for cardiac disease classification (6, 7). ML based computer aided decision support systems enhance diagnostic accuracy and optimize treatment strategies. However, challenges persist in feature selection and model efficiency, as medical datasets often contain redundant and inconsistent information, affecting predictive performance (8). Selecting the most relevant features is crucial for enhancing model accuracy, reducing overfitting, and improving interpretability (9).

Feature Selection methods eliminate redundant data while retaining critical features, improving classification accuracy. RST, introduced by Pawlak, is a powerful mathematical tool for handling uncertainty and incomplete data. It identifies minimal reduct sets while preserving classification accuracy, making it highly suitable for feature selection in ML models (10, 11). MCDM techniques aid in optimizing complex decision processes by distinguishing the best and worst alternatives based on multiple criteria. Decision-making frameworks, such as Fuzzy Set Theory (FST) and RST, effectively handle uncertainty and enhance classification accuracy (12, 13).

Despite advances in ML-based HD prediction, challenges remain in optimizing accuracy and minimizing model complexity. This study addresses these challenges by integrating RST with ML and employing an MCDM-based ranking system to enhance model performance and interpretability. RST with Johnson's algorithm reduces dimensionality by selecting the most informative features, improving both accuracy and interpretability; nine ML classifiers are evaluated and combined into five stacked ensembles through correlation analysis; and MCDM techniques, specifically TOPSIS with MEREC-based weighting, rank the stacked models to identify the best-performing approach, which is further optimized using GridSearchCV. Unlike conventional models that rely on individual classifiers or simple ensembles, this framework enhances accuracy, minimizes overfitting, and improves explainability through Shapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). By

integrating RST, diverse ML classifiers, and MCDM-based ranking, this study addresses key challenges in HD prediction and presents a scalable, robust framework with strong potential for clinical application in early disease detection. The key contributions of the study include,

- Employed RST to effectively eliminate redundant features and retain only the most relevant attributes, thereby enhancing the model's predictive accuracy while minimizing unnecessary complexity. This approach significantly improves HD diagnosis by focusing on essential features.
- Utilized correlation analysis to identify diverse base learners, constructing effective stacked ensemble models. This ensures a robust and versatile prediction model that can handle varied patterns in HD data, leading to enhanced model performance.
- Applied the TOPSIS technique, combined with MEREC-based weighting, to systematically rank the performance of stacked models.
- Fine-tuned Stack-4 using GridSearchCV, ultimately identifying XG as the top-performing classifier. This fine-tuning process led to superior predictive accuracy and ensured the model's robustness in real-world applications.
- Integrated XAI methods such as SHAP and LIME to analyse and interpret feature importance. These techniques enhance the transparency of the model, making it more clinically interpretable and ensuring that healthcare professionals can trust and effectively use the predictive model for early HD diagnosis.

The remainder of the paper is organized as follows: Subsection 1.1 reviews related work, while 1.2 outlines the motivation and highlights the novelty of the proposed approach. Section 2 outlines the materials and methods used in the proposed work. Section 3 presents the research findings. Section 4 discusses the findings and limitations, and Section 5 concludes the study and outlines future work.

1.1 Literature review

Numerous ML methods have been developed recently for the diagnosis of cardiac disease. Furthermore, a variety of medical diagnostics, such as those in Radiology, Dermatology, Hematology, and Ophthalmology, employ AI. HD is the leading cause of death in India, with a study showing a 29.4% prevalence among adults 45 and older. Key risk factors include age, gender, residence, high cholesterol, diabetes, inactivity, and family history, underscoring the need for targeted health programs and early detection efforts (14). Additionally, effective management has significantly improved outcomes for hemophilia patients; however, as they age, they face increased risks of HD, highlighting the imperative for enhanced understanding and management of HD in this population (15).

HD are globally significant, impacting mortality rates and healthcare costs. Timely diagnosis through advanced predictive models like ANN, feature selection methods, and MCDM techniques can effectively reduce fatalities and treatment

expenses. Efficient feature selection and model validation methods highlight promising approaches, offering potential advancements in disease prediction methodologies (16). By providing better risk prediction using AI-based data integration approaches, AI has greatly improved traditional risk assessment tools, such as the Thrombolysis in Myocardial Infarction and Global Registry of Acute Coronary Events scores (17). Decision making techniques, particularly AHP and hybrid approaches, have been prominently used in healthcare contexts to evaluate service quality, offering insights and recommendations for future research directions (18).

Wearable technology monitoring of biopotential signals is critical for HD tracking and early detection. Current developments in wearable biosensors have enhanced precision, reproducibility, and continuous monitoring, significantly reducing healthcare costs (19). Emerging technologies like the FlexiPulse sensor, designed for affordable and accurate HD monitoring, exemplify the advancements in wearable tech. This sensor achieves over 93% accuracy and uses ML to diagnose HD events with 98.7% accuracy (20). Innovative hybrid DL systems have demonstrated high accuracy in predicting HD, utilizing advanced preprocessing techniques and optimized feature extraction methods. For instance, a hybrid DL system achieved 99.12% accuracy across diverse benchmark datasets (21). Furthermore, ML techniques applied to heart failure prediction using advanced methods like XG have shown promising results, achieving accuracies up to 86.36%, highlighting their potential to enhance early mortality estimation and public health outcomes (22).

Furthermore, the analysis of HD has been improved by data mining techniques such as Recurrent Neural Network - Long Short-Term Memory models and genetic algorithms, which demonstrate progress if applied to clinical datasets (23). Novel ML methods, such as those optimized with N2Genetic optimizer, have shown superior accuracy and F1-scores in predicting coronary artery disease, emphasizing their utility in medical decision making. Use of ML in clinical practice extends from pre-clinical data processing to bedside applications, although challenges like validation in real-life settings and ethical considerations remain critical (24). Identification of those with a higher risk of early atherosclerosis and HD in adolescents with Major Depressive Disorder can facilitate individualized therapies and improve outcomes related to both cardiovascular and mental health (25). Using ML to precisely predict HD is essential for early detection and lowering death rates. The study compares several ML algorithms, identifying Random Forest (RF) as the most accurate, and underscores the potential of these technologies to improve healthcare outcomes (26). ML techniques applied to RNA-seq data identified significant genes linked to HD, enhancing early prediction capabilities. The study underscores the potential for these approaches in advancing personalized treatments and understanding disease heterogeneity (27).

In a large population-based study, ML survival models, developed from self-reported questionnaire data, slightly outperformed traditional classification methods in predicting hospitalization for ischemic HD and cardiovascular mortality, with logistic regression being the top-performing classification method. These models demonstrate promise as reliable tools for

screening and identifying individuals at high risk (28). Recent healthcare research emphasizes the importance of early detection of HD, focusing on assessing the Right Ventricle through MRI image segmentation using ML and DL methods such as Fourier-Convolutional Neural Networks (F-CNN), Vanilla CNN, and ResNet. These techniques aim to improve accuracy in identifying the right ventricle abnormalities and enhancing decision making in HD treatment (29). Techniques like disease comorbidity network-based temporal DL framework, which integrates disease comorbidity networks and time-aware DL, enhance cardiovascular risk prediction, especially in patients with mental disorders (30).

Predictive models based on ML for assessing the risk of HD are designed to perform consistently across various demographic categories, such as gender and race. Techniques for mitigating bias are crucial to address systematic biases in health data collection and preprocessing, which can impact model performance on different demographic sub-cohorts. These models use electronic health records data and various ML algorithms to ensure fairness and accuracy in HD risk prediction across diverse populations (31). ML models, particularly RF with Synthetic Minority Oversampling Technique (SMOTE), achieve high accuracy (96.6%), sensitivity (90%), and specificity (100%) in early prediction of cardiac disease using UCI heart dataset (32). A novel multi-modal approach integrates ECG and PCG features to predict HDs. Genetic algorithms optimize feature subsets, and Support Vector Machine (SVM) classification achieves an AUC of 0.936 (33).

ML plays a vital role in healthcare by leveraging extensive datasets to predict HDs early, ensuring better detection and treatment outcomes. Fusion models combining outputs from multiple algorithms achieve high accuracy, reaching 93%–95% in binary classification and 72%–75% in multiclass classification of HD severity. This highlights their potential for early disease identification and risk assessment in healthcare settings (34). The “Sathvi” dataset, compiled from prevalent HD datasets, has boosted prediction accuracy using ML classifiers like CatBoost, achieving up to 98.11% accuracy through Cross-Validation (CV) (35). Various ML algorithms applied to clinical and pathological data, such as Gradient Boosting and XG, have shown that categorical features consistently outperform numerical and combined features, with SVM and AdaBoost achieving optimal performance in CVD prediction (36). Predicting the risks of CVD accurately during pregnancy is crucial for efficient treatment. A hybrid system that combines the Sugeno Fuzzy Inference System (S-FIS), Coefficient of Variation, and Fuzzy Analytic Hierarchy Process (F-AHP) demonstrated high accuracy, sensitivity, and precision, showing its potential in clinical settings for predicting CVD during pregnancy (37).

Unlike prior models that rely heavily on individual classifiers or lack interpretability, the proposed framework integrates RST for efficient feature reduction, stacked ensemble learning for enhanced robustness, and XAI techniques for improved transparency. This holistic approach not only boosts predictive performance but also supports clinical decision-making through interpretability. Additionally, Table 6 provides a detailed

quantitative comparison of our model against existing state-of-the-art HD prediction methods, further justifying the proposed methodological choices.

1.2 Motivation and novelty

Traditional methods for predicting HD often face challenges in handling large volumes of data, high-dimensionality, and uncertainty in feature selection. The motivation for the proposed work arises from the need to improve prediction accuracy by addressing these challenges by combining RST for feature selection with ML techniques, and utilizing MCDM methods for model ranking. The goal is to develop a robust framework that enhances prediction accuracy by selecting the best-performing model through RST-based feature selection and MCDM ranking for HD prediction. The novelty of the proposed work lies in the integration of advanced methodologies to address persistent challenges in ML-based HD prediction:

1. The proposed work applies RST using Johnson's algorithm to reduce data dimensionality by identifying the most informative features. This process enhances model accuracy and interpretability by eliminating redundant attributes.
2. The framework evaluates nine diverse ML classifiers to determine the most effective algorithms for HD prediction. These classifiers are combined into five stacked ensembles through correlation analysis.
3. The incorporation of MCDM techniques, specifically TOPSIS with MEREC-based weighting, systematically ranks the stacked models. This approach identifies the best-performing stack for HD prediction, which is further optimized using GridSearchCV to determine the most effective ML model.

Unlike traditional models that rely on individual classifiers or simple ensembles, this integrated approach enhances prediction accuracy, minimizes overfitting, and offers clearer insights into feature importance through XAI techniques such as SHAP and LIME. By integrating RST, diverse ML classifiers, and MCDM-based ranking, the proposed work fills a critical research gap and presents a scalable, robust framework with strong potential for real-world clinical implementation in early HD detection.

2 Materials and methods

This section provides a comprehensive overview of the methodology employed in this study, detailing the dataset used, preprocessing steps, feature selection techniques, classifier selection, MCDM ranking, and hyperparameter tuning.

2.1 Proposed methodology

This section proposes a hybrid intelligent Optimized RST-ML approach with MCDM-Based Ranking for the diagnosis of HD. A standardized benchmark HD dataset (I_{HD}) from Kaggle (38) was used to evaluate the efficiency of the proposed work. The

dataset comprises features extracted from multiple sources, including data from Switzerland (S_d), Hungary (H_y), Cleveland (C_v), and V Long Beach (L_b) as shown in Equation (1).

$$I_{HD} = \cup(C_v, H_y, S_d, L_b) \quad (1)$$

RST analyses the large features set (I_{HD}) using Johnson's algorithm to generate the minimal subset of attributes (O_{RHD}) to enhance the classification ability which is shown in Equation (2)

$$O_{RHD} = f_{RST}(I_{HD}) \quad (2)$$

The selected features (O_{HD}) are pre-processed with StandardScaler, resulting in the scaled features as shown in Equation (3). This process ensures consistent feature scaling and mitigates the impact of outliers.

$$O_{Rpp} = f_{pp}(O_{HD}) \quad (3)$$

The optimal pre-processed features (O_{Rpp}) are classified using various ML classifiers O_{CL} including XG, AdaBoost (AB), Logistic Regression (LR), K- Nearest Neighbor (KNN), SVM, Naïve Bayes (NB), Decision Tree (DT), RF, and Extra Trees (ET). This classification process is described in Equation (4).

$$O_{CL} = f_{CL}(LR, KNN, SVM, NB, DT, RF, XG, AB, ET) \quad (4)$$

The classification results are evaluated using metrics such as Sensitivity (Sn), F1-score (F1), Specificity (Sp), MCC, MAE, accuracy (Ac), precision (Pr), and training time (Tt). These metrics are represented by O_{MHD} and are given in Equation (5).

$$O_{HD} = O_{Rpp} (O_{CL} (Metrics_{Ac, Pr, Sp, Sn, F1, MCC, MAE, Tt})) \quad (5)$$

To improve the overall performance and robustness of the predictive model, different stacks (O_{s_i}) based on correlation analysis (O_{CA}) has been proposed as shown in Equation (6). These stacks O_{s_i} includes multiple models, resulting in improved accuracy, reduced overfitting, increased flexibility, enhanced stability, and reduced model bias. These stacks O_{s_i} will be used to evaluate the classification performance of CD.

$$O_{s_i} = f_{s_i}(O_{MHD}) \text{ where } i \in 1 \text{ to } 5 \quad (6)$$

MCDM offers a robust framework for addressing complex decision-making scenarios involving multiple classification accuracies obtained by the different stacked models O_{s_i} , often characterized by conflicting criteria. By integrating diverse criteria into the decision-making process, MCDM ensures that decisions are balanced, transparent, and justifiable. In this context, TOPSIS is employed to evaluate and rank the stacks O_{s_i} , facilitating the identification of the optimal stack (O_{BS}) for the

accurate prediction of HD, as shown in Equation (7).

$$O_{BS} = f_{MCDM}(O_{s_i}) \quad (7)$$

The classifiers in the optimal stack O_{BS} , are refined using GridSearchCV to optimize hyperparameters by exploring various parameter combinations, ultimately identifying the best classifier (O_{BC}) for predicting HD. The corresponding equation is provided in Equation (8). The architecture of the proposed optimized RST-ML approach is illustrated in Figure 1.

$$O_{BC} = f_{Op}(O_{BS}) \quad (8)$$

2.2 Dataset description

The Kaggle HD dataset (I_{HD}), sourced from <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>, was used in this investigation (38). This dataset aggregates records from four medical institutions: S_d , H_y , C_v , and L_b . Of the original 76 attributes, a standardized subset of 14 clinically relevant features and a final subset of 303 complete records was selected, primarily from the Cleveland database, which contains no missing values, as adopted in most published studies, to ensure comparability, yielding. This refined dataset improves interpretability and robustness for model training and evaluation. The output label classifies patients into two categories: 1 indicates the presence of cardiac disease, and 0 indicates its

absence. These diagnostic labels were determined through clinical evaluation methods, including electrocardiographic results and stress testing. The selected features encompass critical clinical indicators such as age, blood pressure, cholesterol levels, chest pain type, and other parameters commonly used in cardiac risk prediction. A detailed description of these attributes in the HD dataset is provided in Table 1.

2.3 Johnson's algorithm

Numerous applications of RST have shown it to be effective, including data reduction, identifying hidden trends in data, assessing the significance of data, and creating sets of decision rules based on data (39). To identify the most significant features for analysis, Johnson's Algorithm of RST is used. The Johnson's Algorithm also known as Johnson Reducer, uses a simplified greedy method to find a single reduct R . Johnson (40) pointed out that this approach typically finds a minimal-length prime implicant. Until an ideal subset, or reduct, is found, the algorithm iteratively chooses the most important feature that enhances object distinguishability and updates the attribute set. The group of sets associated with the discernibility function is represented by Q , and the weight assigned to each set Q in Q is provided by $w(Q)$, which is automatically calculated from the data.

RST was chosen over modern feature selection methods for its ability to handle uncertainty and imprecision without prior assumptions or parameter tuning. It identifies minimal subsets of attributes based on data dependencies, offering high interpretability and transparency which is critical in healthcare

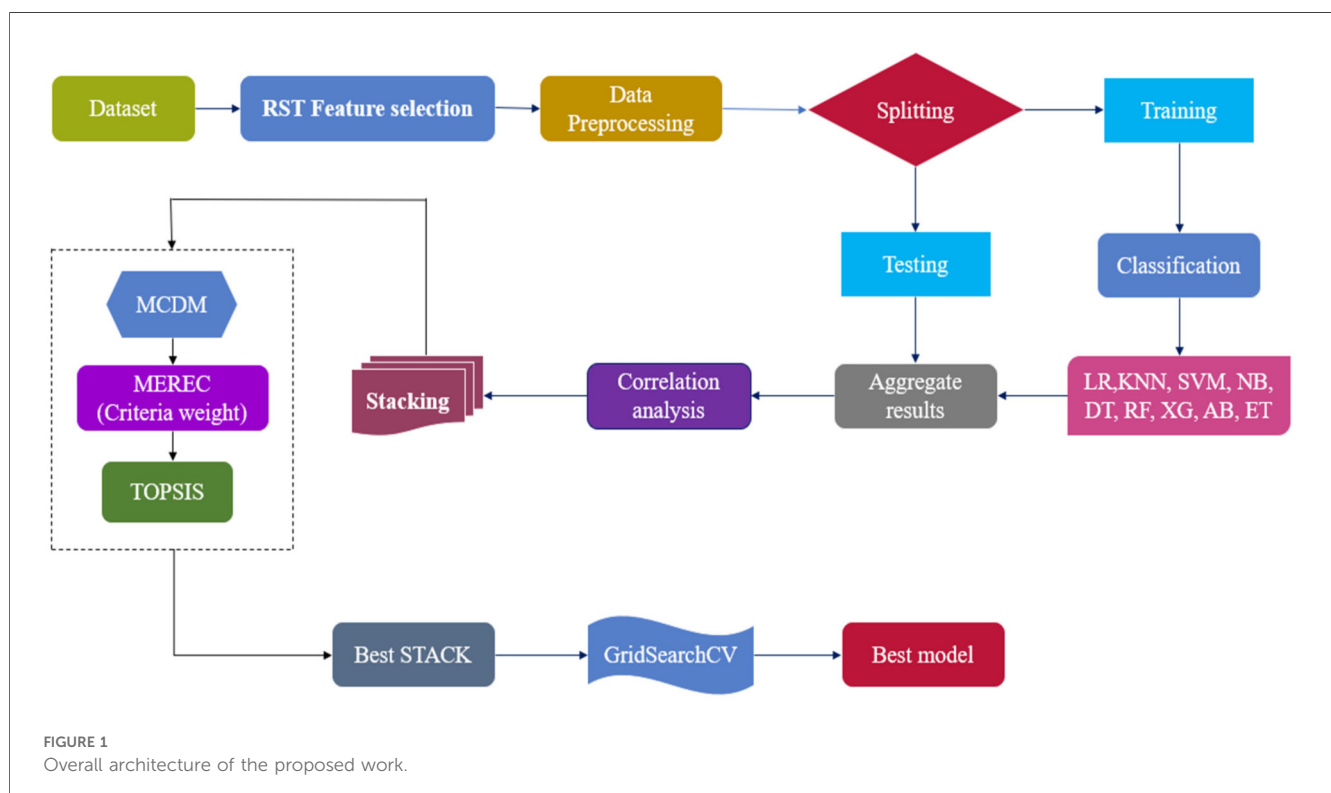


TABLE 1 Detailed attribute description of the heart disease dataset.

Feature	Description	Range	Type
Age	Patient's age in completed years.	29–77	Numeric
Sex	Patient's gender (male = 1, female = 0).	0, 1	Nominal
Cp	Type of chest pain: 1 = Typical angina, 2 = Atypical angina, 3 = Non-anginal pain, 4 = Asymptomatic.	1–4	Nominal
Trestbps	Resting blood pressure in mmHg at the time of admission to the hospital.	94–200	Numeric
Chol	Serum cholesterol in mg/dl.	126–564	Numeric
Fbs	Fasting blood sugar > 120 mg/dl (1 = Yes, 0 = No).	0, 1	Nominal
Restecg	Resting electrocardiogram results: 0 = Normal, 1 = ST-T wave abnormality (e.g., T-wave inversion or ST elevation/depression > 0.05 mV), 2 = Left ventricular hypertrophy by Estes' criteria.	0–2	Nominal
Thalach	Maximum heart rate achieved.	71–202	Numeric
Exang	Exercise-induced angina (0 = No, 1 = Yes)	0, 1	Nominal
Oldpeak	ST depression induced by exercise relative to rest.	0–6.2	Numeric
Slope	Slope of the peak exercise ST segment: 1 = Upsloping, 2 = Flat, 3 = Downsloping.	1–3	Nominal
Ca	Number of major vessels (0–3) colored by fluoroscopy.	0–4	Numeric
Thal	Status of the heart: 3 = Normal, 6 = Fixed defect, 7 = Reversible defect.	3, 6, 7	Nominal
Target	HD diagnosis: 0 = No disease, 1 = Presence of disease	0–1	Nominal

applications. Compared to statistical or black-box techniques like Lasso or embedded models, RST provides a rule-based, explainable framework while preserving classification accuracy during dimensionality reduction. This makes RST, and Johnson's Algorithm in particular, a robust and suitable choice for the proposed HD prediction model. Algorithm 1 outlines the steps involved in Johnson's reducer of RST.

2.4 Dataset preprocessing

The significant features O_{RHD} , obtained from Johnson's Algorithm of RST was preprocessed using the StandardScaler technique to ensure uniformity in feature scales and improve

model training. StandardScaler standardizes data by subtracting the mean and dividing by the standard deviation, as shown in

$$z = \left(\frac{x - \bar{x}}{s} \right) \quad (9)$$

where x is the original feature value, $\bar{x}$ is the mean, and s is the standard deviation.

This preprocessing step ensures that all features contribute equally to the model, eliminating biases caused by differences in scale. By enhancing convergence during training and mitigating the influence of outliers, StandardScaler improves model performance, resulting in more robust and reliable predictions. The dataset was divided into an 80:20 train-test split ratio, with 80% of the data allocated for training the model and 20% reserved for testing its performance to evaluate and identify the most effective model.

2.5 Machine learning models

This section highlights various ML algorithms, each chosen for its unique ability to process diverse data types. ML models have become essential tool for predicting HD due to their capability to analyse complex datasets and uncover valuable insights.

LR offers simplicity and interpretability for binary classification (41) but assumes linear relationships, which can limit its performance on complex data. KNN, a non-parametric method based on proximity (42), is intuitive but sensitive to noise and suffers in high-dimensional spaces. SVM provides robust classification in high-dimensional settings (43), yet is computationally intensive and requires careful kernel tuning. NB, though efficient (44), relies on the unrealistic assumption of feature independence. DT, valued for their transparency (45), are prone to overfitting; hence, ensemble approaches were used to improve generalization. These limitations were addressed using feature scaling, selection, and cross-validation.

To enhance performance and mitigate overfitting, ensemble models including RF, AB, XG, and ET were incorporated. RF reduces variance through bagging (46) but at the cost of interpretability. XG, known for its high accuracy, incorporates L1/L2 regularization (47) but requires extensive tuning. AB adapts to misclassified instances for improved robustness (48), though it is sensitive to noisy labels. ET increases diversity via random splits (49), but excessive randomness may cause instability in small datasets. Each model contributed unique strengths across varying data complexities, allowing a comprehensive evaluation of HD risk prediction (50, 51).

2.6 Classifier selection and stacking strategy

The classifiers for stacking were chosen based on Pearson correlation analysis to ensure diversity and complementarity. Five distinct stacks were created, each with specific combinations of classifiers designed to optimize performance. Stack-1 (Hybrid Stack) combines LR, KNN, and SVM, which are highly correlated and offer similar performance. Stack-2 (Tree and Probabilistic Stack) integrates NB, DT, and RF, with NB adding

Algorithm 1 Johnson's Algorithm for finding reducts RST.

Input: Information table or Decision Table, $I = (U, A)$
Output: $RED_A \subseteq A$ of all reducts of A
Step 1: Initialize the reduct set R as an empty set $R = \emptyset$
Step 2: Find the attribute a that maximizes the sum of weights $\sum w(Q)$, where the sum is taken over all sets Q in $\mathcal{Q}$ that contain a . $a = \arg \max_{a \in \text{Attributes}} \sum_{\substack{Q \in \mathcal{Q} \\ a \in Q}} w(Q)$
Step 3: Add attribute a to R to maximize the weight sum $R = R \cup \{a\}$
Step 4: Remove all sets Q from $\mathcal{Q}$ that contain a to avoid redundant calculation. $\mathcal{Q} = \mathcal{Q} - \{Q \mid a \in Q\}$
Step 5: If $\mathcal{Q} = \emptyset$, return R . If not, return to step 2.

diversity due to its lower correlation with tree models. Stack-3 (Boosted Stack) includes XG, AB, and ET, which are highly correlated to strengthen combined effect, with AB providing unique performance benefits. Stack-4 (Advanced Tree-Based Stack) leverages the high correlations among DT, RF, XG, AB, and ET, maintaining robustness. Stack-5 (Unified Stack) incorporates all nine models to cover a wide range of correlations and performance traits. This stacking approach enhances performance by combining multiple classifiers, allowing the meta-model to improve predictions by learning which base models are most reliable in specific scenarios. Correlation analysis ensures the base model's errors are sufficiently diverse, maximizing the benefits of stacking. The strategy significantly improves prediction accuracy and generalization by leveraging the complementary strengths of various classifiers, thus enhancing the overall model's robustness.

2.6.1 Correlation analysis

Correlation analysis, denoted as O_{CA} , measures the linear relationship between predictions from different classifiers to ensure diversity in stacking. The proposed work utilizes Pearson correlation to identify and group classifiers based on their performance similarity, aiming to balance diversity and complementarity in each O_{S_i} for optimal performance. The procedure for correlation analysis is detailed in [Algorithm 2](#).

2.6.2 Stacked classifier

Stacking, or stacked classifiers ([52](#)), is an ensemble learning technique that combines several Base Classifiers (BC) through a Meta-Classifier (MC) to improve performance. This approach

trains multiple models on the same data and uses their predictions as inputs for a meta-model, which leverages the strengths of each model while mitigating its weaknesses. The principle is that no single algorithm is optimal for all problems. If $\{h_1, h_2, \dots, h_n\}$ are base classifiers and H is the meta-classifier, the final prediction $\hat{y}$ is given by Equation (10)

$$H(h_1(x), h_2(x), \dots, h_n(x)) \quad (10)$$

where x is the input feature vector.

In the proposed work, different stacks O_{S_i} is employed to identify the best prediction model for predicting HD. [Figure 2](#) illustrates each phase of the O_{S_i} , detailing the process from base model training to final predictions.

The [Algorithm 3](#) provided below outlines the sequential steps involved in combining the predictions of multiple base models to produce the final output.

2.7 MCDM ranking

To refine the selection of the best-performing stack O_{BS} , a MCDM approach is implemented, addressing the challenge of repeated accuracy values across stacked classifiers O_{S_i} . By incorporating multiple evaluation metrics including Sn , $F1$, Sp , MCC , MAE , Ac , Pr and Tt , a comprehensive and balanced assessment is ensured. MCDM enables a detailed comparison across these criteria, fostering robust and well-informed decision-making.

Although AUC is a commonly used and intuitive metric in healthcare for measuring a model's ability to discriminate between classes, it captures only one dimension of performance. The proposed work employs MCDM ranking to evaluate models across multiple metrics simultaneously, including Acc , Pr , Rc , $F1$, MCC , MAE and Tt . This approach provides a comprehensive evaluation framework that accounts for various aspects of model performance, reflecting the complex trade-offs often encountered in clinical decision-making. For example, a model with a high AUC might have suboptimal recall or precision, which could have important implications in patient diagnosis or treatment.

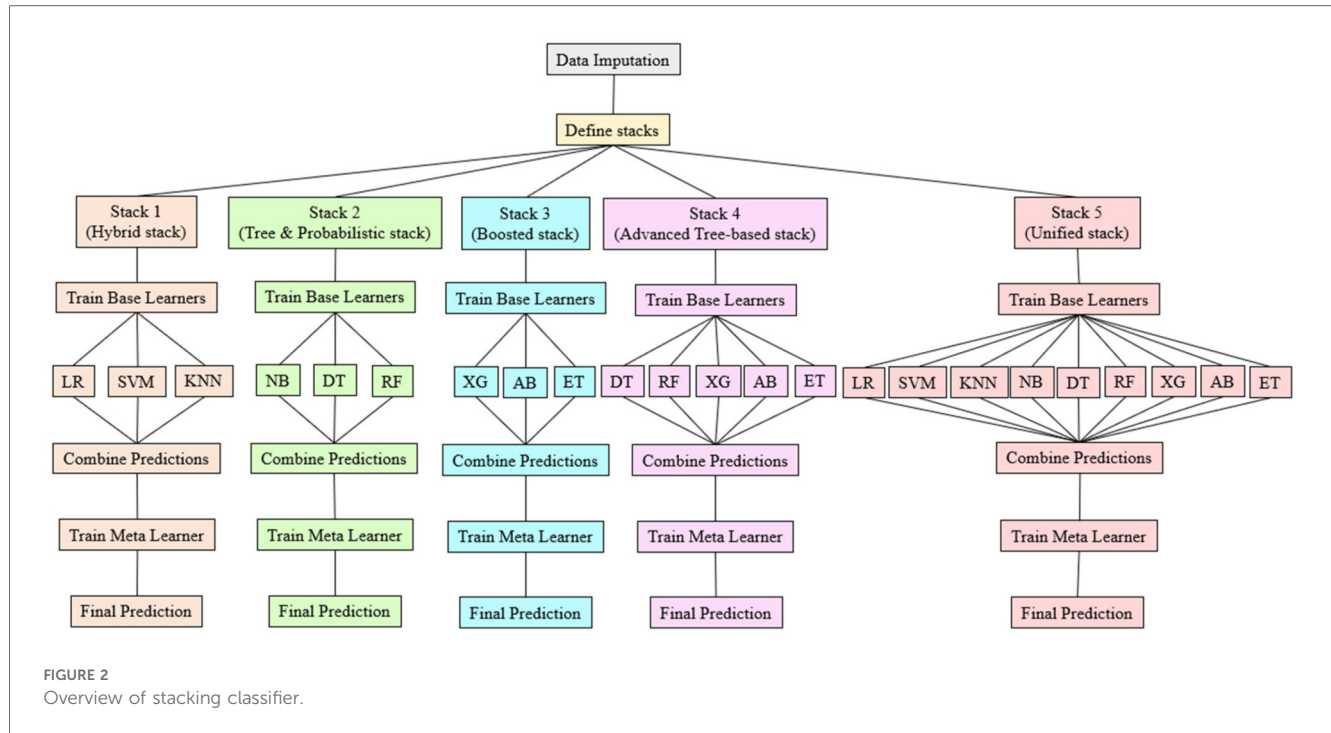
To facilitate understanding among healthcare professionals, the MCDM rankings was presented alongside detailed explanations of each contributing metric, thereby ensuring transparent and actionable insights that extend beyond the scope of AUC alone. Specifically, the TOPSIS method is used for ranking the stacks, and the MEREC method is employed for objective criteria weighting.

2.7.1 MEREC weighting method

The MEREC method, introduced by Keshavarz-Ghorabae et al., calculates criteria weights by evaluating the influence of each criterion on the performance of alternatives ([53](#)). This method assigns higher weights to criteria with a more significant impact, ensuring an objective weighting process. The steps of MEREC are outlined in [Algorithm 4](#):

Algorithm 2 Grouping ML models as Stacks based on O_{CA} .

Input: RST – ML metrics (O_{MHD})
Output: Different stacks (O_{S_i})
Step 1: Imputation of Performance metrics as matrix $X = M_{ij}$, where $1 \leq i \leq 9$ & $1 \leq j \leq 8$ where i represents model and j represents performance metrics.
Step 2: Compute Correlation Coefficients
Step 2.1: Covariance Calculation $Cov(X_i, X_j) = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{n - 1}$
Step 2.2: Variance Calculation $Var(X_i) = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)^2}{n - 1}$ $Var(X_j) = \frac{\sum_{k=1}^n (X_{jk} - \bar{X}_j)^2}{n - 1}$
Step 2.3: Pearson Correlation Coefficient $r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i) \cdot Var(X_j)}}$
Step 3: Compute Correlation Matrix $CM = C_{ij}$, where $1 \leq i \leq 8$ & $1 \leq j \leq 8$
Step 3.1: Interpretation Step 3.1 (a): High Positive Correlation ($0 < c \leq 1$) Step 3.1 (b): Zero Correlation ($c = 0$) Step 3.1 (c): Negative Correlation ($-1 \leq r < 0$)
Step 4: Group Classifiers into stacks based on Correlation O_{S_i} , where $1 \leq i \leq 5$



Algorithm 3 Performance evaluation of Stacking classifiers.

Input: RST preprocessed dataset (O_{Rpp})
Output: Performance of Stacked classifiers ($O_{S_{MHD}}$)

Step 1: Data Imputation
 $X = O_{Rpp}$

Step 2: Split Dataset
 $X_S = (X_{train}, X_{test}, y_{train}, y_{test}) = S(X, y, 0.2, RS = 42)$

Step 3: Define Stacks

Step 3.1: Stack 1 – Hybrid Stack
 $BC_1 = \{LR_{O_{Rpp}}, KNN_{O_{Rpp}}, SVM_{O_{Rpp}}\}; MC_1 = M_{meta, 1}$

Step 3.2: Stack 2 – Boosted Stack
 $BC_2 = \{NB_{O_{Rpp}}, DT_{O_{Rpp}}, RF_{O_{Rpp}}\}; MC_2 = M_{meta, 2}$

Step 3.3: Stack 3 – Tree and Probabilistic stack
 $BC_3 = \{XGO_{Rpp}, ABO_{Rpp}, ETO_{Rpp}\}; MC_3 = M_{meta, 3}$

Step 3.4: Stack 4 – Advanced tree-based stack
 $BC_4 = \{DT_{O_{Rpp}}, RF_{O_{Rpp}}, XGO_{Rpp}, ABO_{Rpp}, ETO_{Rpp}\}; MC_4 = M_{meta, 4}$

Step 3.5: Stack 5 – Unified stack
 $BC_5 = \{LR_{O_{Rpp}}, KNN_{O_{Rpp}}, SVM_{O_{Rpp}}, NB_{O_{Rpp}}, DT_{O_{Rpp}}, RF_{O_{Rpp}}, XGO_{Rpp}, ABO_{Rpp}, ETO_{Rpp}\};$
 $MC_5 = M_{meta, 5}$

Step 4: Train and evaluate each stack S_i ($i = 1$ to 5)

Step 4.1: BC
 $\hat{Y}_{train, i} = [CL_{ij}(X_{train}) \text{ for } j \text{ in } BC_i]; \hat{Y}_{test, i} = [CL_{ij}(X_{test}) \text{ for } j \text{ in } BC_i]$

Step 4.2: Combine predictions from BC
 $X_{meta, train, i} = \hat{Y}_{train, i}; X_{meta, test, i} = \hat{Y}_{test, i}$

Step 4.3: Train MC
 $M_{meta, i} \cdot fit(X_{meta, train, i}, y_{train, i})$

Step 4.4: Predict using MC
 $\hat{Y}_{meta, i} = M_{meta, i} \cdot predict(X_{meta, train, i})$

Step 5: Performance evaluation
 $S_{iMHD} = f_{Y_{meta, i}}(Metrics_{Acc, Pr, Rc, F1, MCC, MAE, T1})$

Algorithm 4 MCDM Criteria weight calculation.

Input: S_{iMHD} as Decision matrix (dm)
Output: Criteria weight (w_r)

Step 1: Establish the Principal dm
 $dm = [s_{tr}] = \begin{bmatrix} s_{11} & \dots & s_{1r} \\ \dots & \dots & \dots \\ s_{r1} & \dots & s_{rr} \end{bmatrix}, \text{ where } 1 \leq t \leq m,$
 $1 \leq r \leq n$

Step 2: Normalization of dm
 $N_{tr} = \frac{\min x_{tr}}{x_{tr}}, \text{ if } r \text{ is a beniteficial creria}$
 $N_{tr} = \frac{x_{tr}}{\max x_{tr}}, \text{ if } r \text{ is a non - beniteficial creria}$

Step 3: Determine the Alternative's Overall Performance (P_t)
 $P_t = \ln \left(1 + \left(\frac{1}{m} \sum_r |\ln(n_{tr})| \right) \right)$

Step 4: Re-evaluate the performance of each P_t by excluding the impact of each criterion.
 $P'_{tr} = \ln \left(1 + \left(\frac{1}{m} \sum_{l \neq r} |\ln(n_{tr})| \right) \right)$

Step 5: Calculate Absolute Deviations (Ad)
 $Ad_r = \sum_t |P'_{tr} - P_t|$

Step 6: Determine Criteria Weights (w_r)
 $w_r = \frac{Ad_r}{\sum_r Ad_r}$

2.7.2 TOPSIS alternative assessment

Once the criteria weights (w_r) are determined using MEREC, the stacks O_{S_i} are ranked using the TOPSIS method. Developed by Hwang and Yoon in 1981, TOPSIS evaluates alternatives based on their proximity to an ideal solution (positive ideal) and

Algorithm 5 Ranking Alternatives based on TOPSIS.

Input: S_{imdd} as Decision matrix (dm)**Output:** Rank r **Step 1:** Establish dm

$$dm = [s_{tr}] = \begin{bmatrix} s_{11} & \dots & s_{1r} \\ \dots & \dots & \dots \\ s_{t1} & \dots & s_{tr} \end{bmatrix}, \text{ where } 1 \leq t \leq m, \\ 1 \leq r \leq n$$

Step 2: Normalizing dm (Ndm) with Vector Normalization

$$r_{tr} = \frac{x_{tr}}{\sqrt{\sum_{t=1}^m x_{tr}^2}}$$

Step 2: Compute the weighted Ndm .

$$u_{tr} = r_{tr} \times w_r, \text{ where } 1 \leq t \leq m, 1 \leq r \leq n \\ U = \begin{bmatrix} u_{11} & \dots & u_{1n} \\ \dots & \dots & \dots \\ u_{m1} & \dots & u_{mn} \end{bmatrix}$$

Step 3: Identify the positive (A^+) and negative (A^-) ideal solutions.

$$A^+ = \{(\max u_{tr} | r \in R), \min v_{tr} | r \in R'\} \\ A^- = \{(\min u_{tr} | r \in R), \max v_{tr} | r \in R'\}$$

Step 4: Calculate separation measures of positive (S_i^+) and negative (S_i^-) ideal solutions.

$$S_i^+ = \sqrt{\sum_{t=1}^m (u_{tr} - u_t^+)^2} \\ S_i^- = \sqrt{\sum_{t=1}^m (u_{tr} - u_t^-)^2}$$

Step 5: Determine the relative closeness from the ideal Solution (C_i)

$$C_i = \frac{S_i^-}{(S_i^- + S_i^+)}$$

Step 6: Rank (r) the alternatives according to preference order.

their distance from a worst-case scenario (negative ideal) (54). This ensures that the chosen Stack O_{BS} exhibits optimal performance across all criteria. The steps of TOPSIS, are outlined in Algorithm 5:

2.8 Gridsearchcv for hyperparameter tuning

To identify best classifier O_{BC} for HD prediction, the top-performing stack O_{BS} , determined through MCDM-TOPSIS ranking, undergoes hyperparameter optimization using GridSearchCV (55). GridSearchCV systematically explores predefined hyperparameter ranges for each classifier within O_{BS} , optimizing parameters such as learning rate, tree depth, and regularization strength. Leveraging CV, it evaluates parameter combinations to balance overfitting and underfitting, ensuring robust model performance. Each classifier in O_{BS} is fine-tuned individually, and the one achieving the highest validation score post-tuning is selected as O_{BC} . This comprehensive process ensures that the best-performing classifier is identified, optimized, and validated, resulting in a reliable and accurate predictive model for HD prediction. The detailed process of ranking and optimization of stacked classifiers using MCDM-TOPSIS and GridSearchCV is provided in Figure 3.

2.9 Statistical Analysis

The model's performance is assessed using various standard metrics, including Ac , Pr , Sn , Sp , $F1$, MCC , MAE and Tt . To compare the performance across different classifiers, statistical tests such as paired t -tests and Wilcoxon signed-rank tests are applied.

3 Research findings

This section presents the key findings of the study, including RST-based feature selection using Johnson's algorithm, evaluations of ML models O_{CL} , performance analysis of stacked models O_{Si} , MCDM-TOPSIS ranking, and GridSearchCV optimization of the top ranked stack O_{BS} .

3.1 Experimental setup

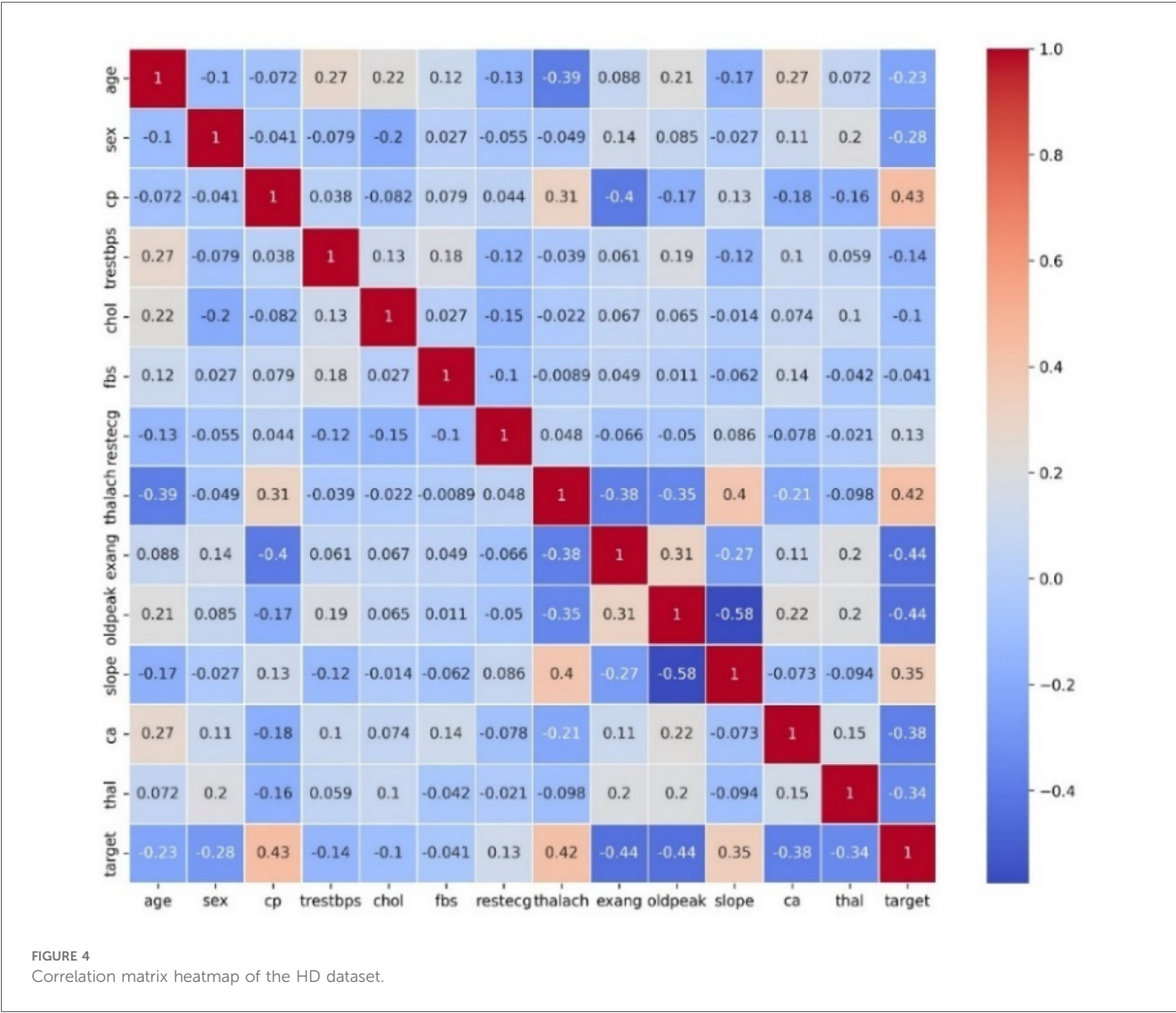
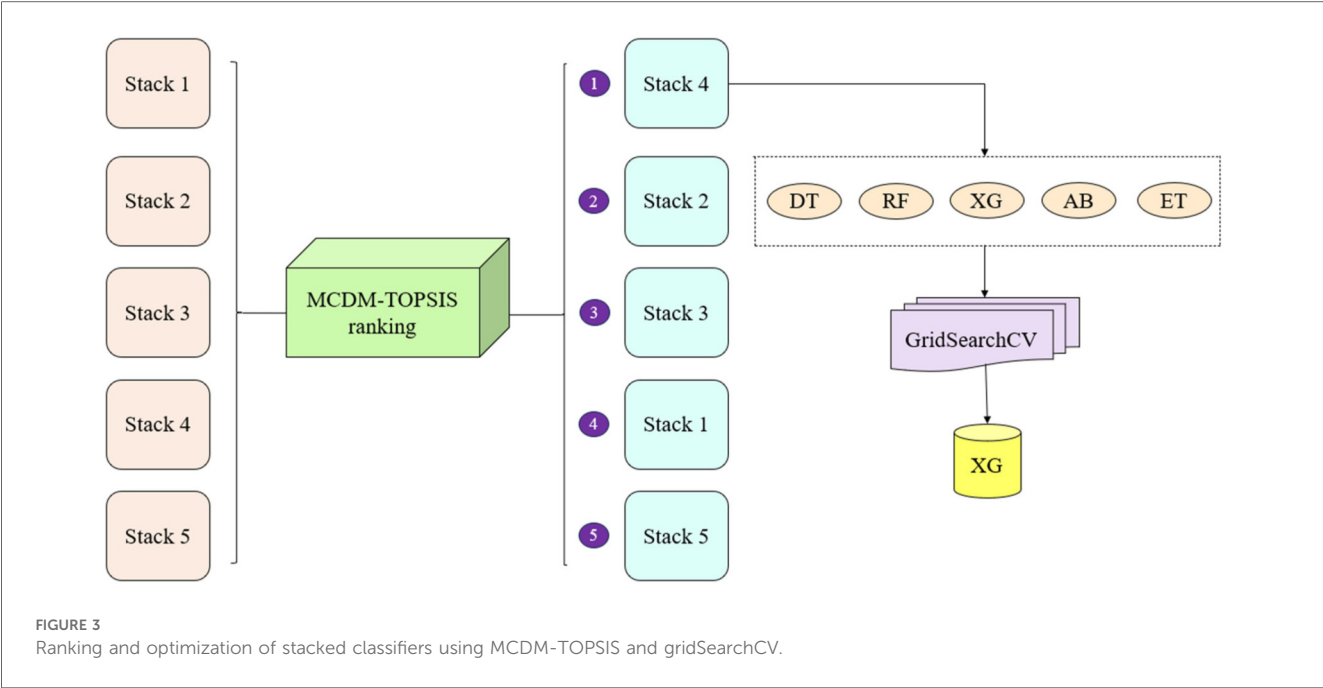
The experiments were conducted using ROSETTA for RST feature selection, Python for ML model development, and MATLAB for MCDM-TOPSIS ranking. The model training and optimization were performed in Google Colab, a cloud-based Jupyter notebook environment, utilizing PyTorch for ML implementation. The system used for training was equipped with a 12th Gen Intel Core i3-1215U processor (1.20 GHz) and 8 GB RAM, ensuring efficient data processing and optimal performance for HD prediction.

3.2 Correlation heat map of HD dataset

Figure 4 displays a heat map which presents a correlation matrix of the features in the HD dataset. The correlations between feature pairs are shown visually in this depiction, where the color intensity corresponds to the direction and strength of the correlation coefficients. This heat map serves as a valuable tool for exploring and understanding the intricate relationships between different features in the dataset, guiding subsequent data preprocessing and modeling steps.

3.3 Feature selection using rough set theory

Feature selection in this study was carried out using RST, which utilizes the concept of reducts to identify the most important features. The aim of this process was to reduce the dimensionality of the dataset while preserving the critical attributes needed to predict HD. To achieve this, Johnson's algorithm is applied within the ROSETTA software, a widely recognized method in RST for identifying the minimal set of features that maintain the decision-making capability of the dataset. The original dataset contained 14 features, but after applying RST-based feature selection, only three features namely, Age, resting blood pressure and serum cholesterol was selected.



This reduction in features simplifies the model, improves computational efficiency, and lowers the risk of overfitting, while maintaining the essential predictive power. These selected features were then used for model training and evaluation, ensuring that the model focused on the most influential variables for accurate HD prediction.

3.4 Model performance evaluation for HD prediction using ML

After selecting the most significant features through RST, the dataset was preprocessed and used to train several ML models, O_{CL} . The performance of these models is presented in Table 2 which provide valuable insights for further model enhancement through techniques like hyperparameter tuning and stacking.

3.5 Enhancing HD prediction through stacking

To improve HD prediction accuracy, stacked ensemble learning was applied. The base learners for these stacking models were carefully selected through Pearson correlation analysis O_{CA} , ensuring a balance between diversity and complementarity. Figure 5 presents the Pearson Correlation Matrix, illustrating performance relationships across models, while Table 3 shows the results of stacking classifiers that combine multiple ML models.

3.6 MCDM –TOPSIS approach

The results from the stacking models in Table 3 shows that Stack-2 to Stack-5 exhibit very similar performance across most metrics, while Stack-1 has lower performance across all metric. The TOPSIS-based MCDM technique was employed to identify the top-performing stack. Table 3 serves as the primary decision for the analysis. Weights were assigned to each criterion using the MEREC method, which measures the impact of removing a criterion on overall performance. This approach ensures that the weight assignment reflects the significance of each criterion in the decision-making process. The criteria weights determined by the MEREC approach are shown in Table 4.

These weights were then incorporated into the weighted normalization process, forming the foundation for the subsequent TOPSIS steps. The results of the TOPSIS analysis, depicted in Figure 6, reveal that Stack-4 stands out as the optimal model, achieving the highest score of 0.8933. This score reflects its superior predictive performance and robustness compared to the other stacks.

3.7 Hyperparameter tuning and performance evaluation of stack 4

To find the best ML classifier, O_{BC} , from Stack – 4, GridSearchCV is utilized to tune hyperparameters. This method systematically explores various combinations and uses CV to evaluate performance. By identifying the optimal hyperparameters, GridSearchCV ensures selection of the best-performing ML model. The results of this comprehensive evaluation are presented in Table 5.

Figure 7 presents a detailed comparison of the metrics across the models in Stack-4, underscoring the effectiveness in identifying the top classifier. It shows that XG is the highest-performing model among other models in Stack 4.

3.8 Statistical and practical evaluation of model performance

To assess the statistical significance of performance differences between the models, paired t -tests and Wilcoxon Signed-Rank tests was conducted. The results revealed no statistically significant differences, with all p -values being greater than 0.05. While trends were observed, such as a p -value of 0.09097 for the comparison between DT and XG, these were not strong enough to reject the null hypothesis. From a practical standpoint, XG excelled in terms of A_c , F_1 , MCC, and MAE. Moreover, XG demonstrated an impressive T_t of just 0.04 s, considerably faster than other models. Although the statistical tests showed no significant differences between the models, the superior performance of XG, coupled with its fast T_t , makes it the ideal choice for this task. Therefore, the final model selection is influenced by performance metrics and the practical considerations such as training efficiency and ease of interpretation, with XG standing out as the optimal model.

TABLE 2 Results of ML models.

S. No.	Model	A_c	P_r	S_p	S_n	F_1	MCC	MAE	T_t
1	LR	0.62	0.65	0.64	0.60	0.63	0.24	0.38	3.28
2	KNN	0.90	0.93	0.92	0.88	0.91	0.81	0.10	0.16
3	SVM	0.61	0.66	0.69	0.54	0.59	0.23	0.39	0.31
4	NB	0.63	0.68	0.69	0.58	0.63	0.27	0.37	0.04
5	DT	0.98	0.98	0.97	0.98	0.98	0.95	0.02	0.37
6	RF	0.98	0.98	0.97	0.98	0.98	0.95	0.02	13.3
7	XG	0.92	0.89	0.87	0.97	0.93	0.85	0.08	2.16
8	AB	0.73	0.74	0.71	0.75	0.75	0.46	0.27	4.56
9	ET	0.98	0.98	0.97	0.98	0.98	0.95	0.02	7.03

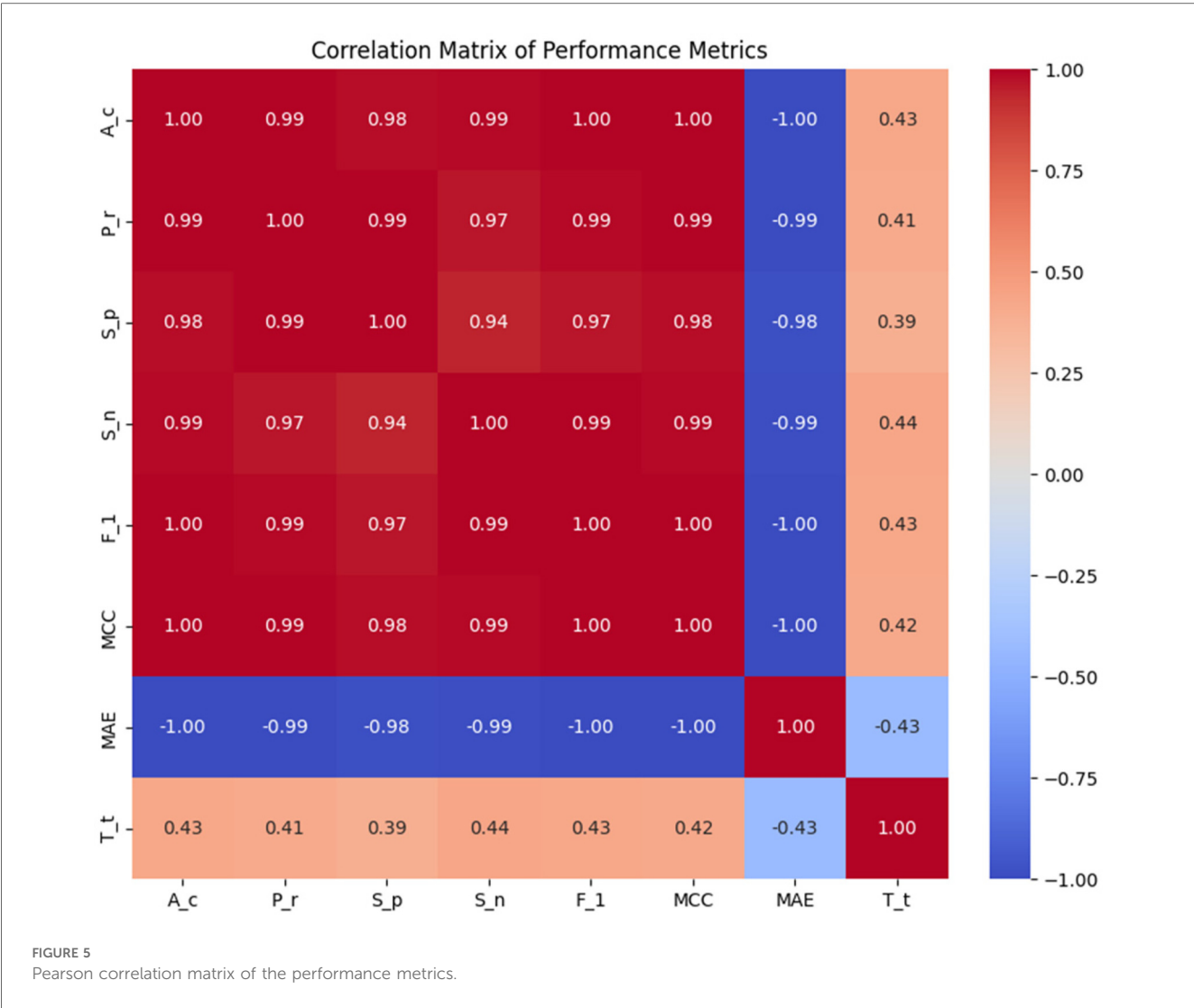


TABLE 3 Performance evaluation of the stacking classifiers.

Model	A_c	P_r	S_p	S_n	F_1	MCC	MAE	T_t
Stack – 1	0.75	0.74	0.79	0.70	0.76	0.49	0.25	0.17
Stack – 2	0.98	0.96	1.00	0.96	0.98	0.96	0.02	1.33
Stack – 3	0.98	0.96	1.00	0.96	0.98	0.96	0.02	1.83
Stack – 4	0.98	0.96	1.00	0.96	0.98	0.96	0.02	0.59
Stack – 5	0.98	0.96	1.00	0.96	0.98	0.96	0.02	3.16

TABLE 4 Criteria Weights determined by the MEREC Method.

Method	A_c	P_r	S_p	S_n	F_1	MCC	MAE	T_t
MEREC	0.0669	0.0683	0.0732	0.0574	0.0695	0.0496	0.2264	0.3685

3.9 Interpretability of the prediction model using SHAP and LIME

After selecting XG as the best model for HD prediction, SHAP and LIME is used for model interpretability to ensure its decisions are based on the right features. SHAP’s summary, dependence, and

force plots highlight key features and their impact on predictions, while LIME’s bar plots provide detailed explanations for individual predictions, especially in edge cases. These techniques help validate the model’s behaviour, ensuring it aligns with domain knowledge and can be trusted for real-world applications.

3.9.1 SHAP interpretation

The SHAP summary plot in Figure 8 provides a concise overview of feature importance across all data points. It ranks features by their influence on the model's predictions for HD risk, with "chol" as the most impactful, followed by "age" and "trestbps". Each point represents a sample, with the x -axis showing the SHAP value, indicating the feature's contribution to risk: positive values increase risk, while negative values decrease it. The color gradient reflects feature values, highlighting trends like higher cholesterol levels being strongly associated with increased risk.

Figure 9 reveals a clear negative correlation between cholesterol levels and their impact on model predictions. As cholesterol values increase along the x -axis, their corresponding SHAP values generally decrease on the y -axis, indicating that higher cholesterol levels contribute more strongly to predicting HD risk. Lower cholesterol values tend to have positive SHAP values, suggesting they decrease HD risk, while higher cholesterol values generally have negative SHAP values, indicating increased risk. The vertical spread of points at each cholesterol level and the variation in colours further emphasizes that the relationship between cholesterol and HD prediction is complex and moderated by other factors in the model.

SHAP force plot explains the contribution of each feature to a single prediction. It highlights how each feature contributes positively or negatively to the model's prediction, allowing users to understand why a particular instance is classified as "high risk" or "low risk" for HD. The force plot shows the aggregated contribution of each feature towards the model's final prediction, which is crucial for interpreting individual predictions. Figure 10

explains a single prediction, showing how the model's base value of 0.5 is adjusted to a final output of -2.65 by three feature contributions. The blue segments represent features lowering the HD risk, with their cumulative effect resulting in a low-risk prediction. This plot provides a clear, quantitative breakdown of how each feature impacts the final prediction.

3.9.2 LIME interpretation

LIME offers a local explanation for a single prediction, showing the importance of features for that specific prediction. In bar plots, features are ranked according to their contribution to the prediction, which can be particularly useful in understanding the reasons behind misclassifications. Figure 11 highlights the contributions of three features to a specific HD prediction. Age has the strongest negative impact (-0.2462), significantly reducing risk, followed by blood pressure with a moderate negative effect (-0.1243). Cholesterol contributes slightly positively ($+0.0786$), marginally increasing risk. The normalized feature ranges and impact magnitudes (-0.25 to 0.05) collectively point to a low-risk classification, offering a clear, interpretable explanation for this individual prediction.

3.10 Performance evaluation of proposed model with state-of-the-art models

A comprehensive comparison of state-of-the-art techniques for HD prediction is presented, highlighting their methodologies, strengths, limitations, and accuracies. These approaches range from traditional ML models to advanced hybrid techniques, with a particular focus on feature selection strategies and dataset utilization. Notably, the proposed Optimized RST-ML model demonstrates superior accuracy, combining RST, ML, stacked ensemble, and MCDM. This hybridized approach, which has not been previously proposed for HD prediction in the literature, offers a novel and robust solution for HD detection and clinical decision-making. The detailed comparison of these techniques is provided in Table 6.

3.11 Proposed model validation with different open source datasets

The proposed Optimized RST-ML approach with MCDM-based ranking was evaluated across multiple open-source datasets to validate its efficiency in HD prediction. The datasets utilized for validation include the Obesity Levels dataset, Breast Cancer

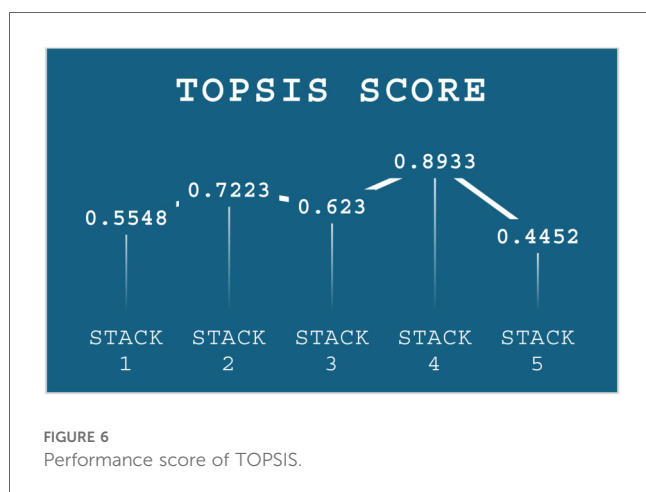


TABLE 5 Performance metrics of ML models in STACK- 4 tuned with gridSearchCV.

Model	A_c	P_r	S_p	S_n	F_1	MCC	MAE	T_t
DT	0.9773	0.9756	0.9816	0.9724	0.9786	0.9544	0.0227	0.003
RF	0.9773	0.9756	0.9816	0.9724	0.9786	0.9544	0.0227	0.19
XG	0.9870	0.9760	1.00	0.9724	0.9879	0.9742	0.0129	0.04
AB	0.7825	0.7892	0.8037	0.7586	0.7964	0.5630	0.2175	4.56
ET	0.9773	0.9756	0.9816	0.9724	0.9786	0.9544	0.0227	7.03

The bold values highlight the best results, obtained by XG, across each evaluation metric.

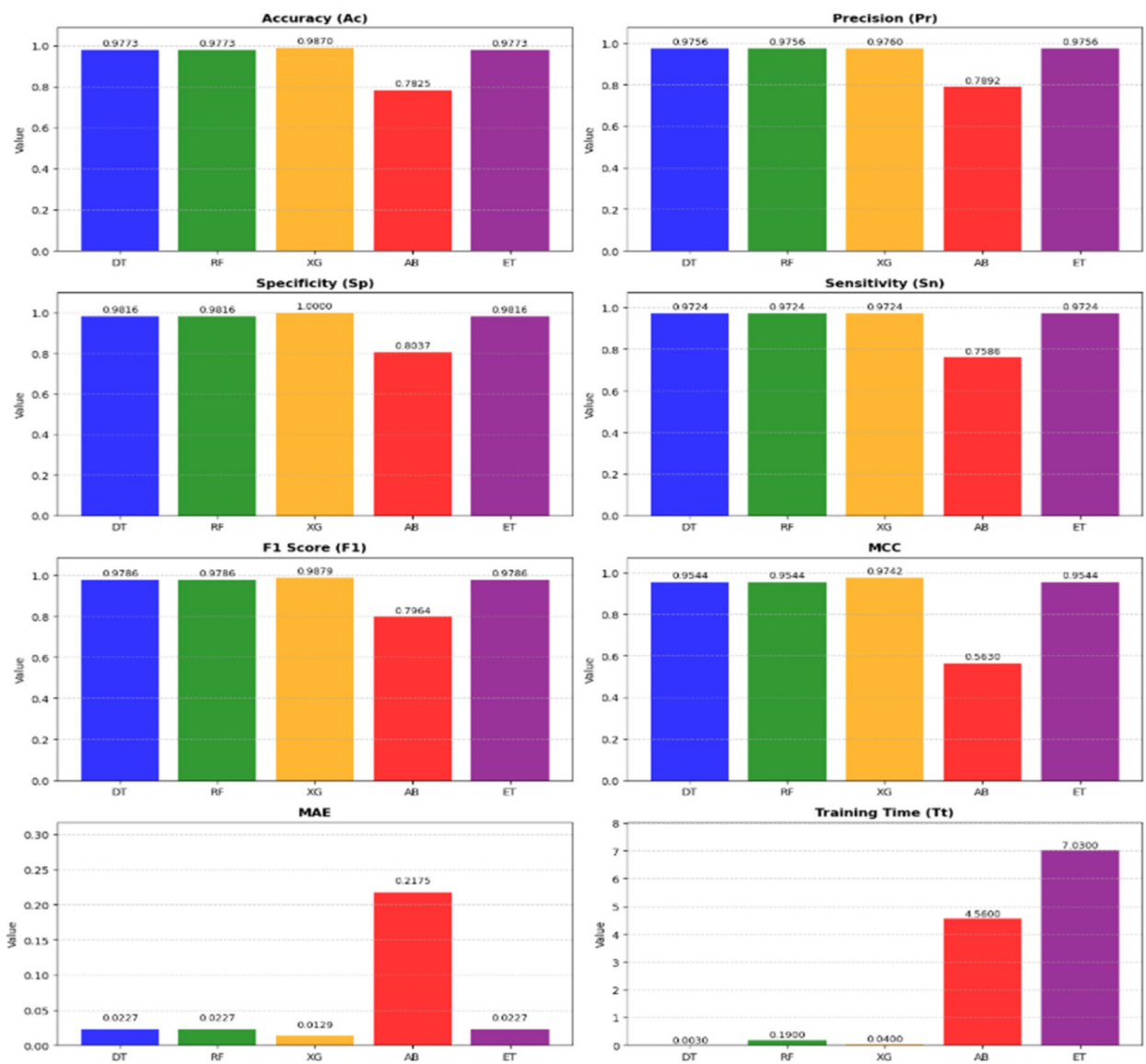


FIGURE 7 Metrics-wise performance comparison of stack-4 models.

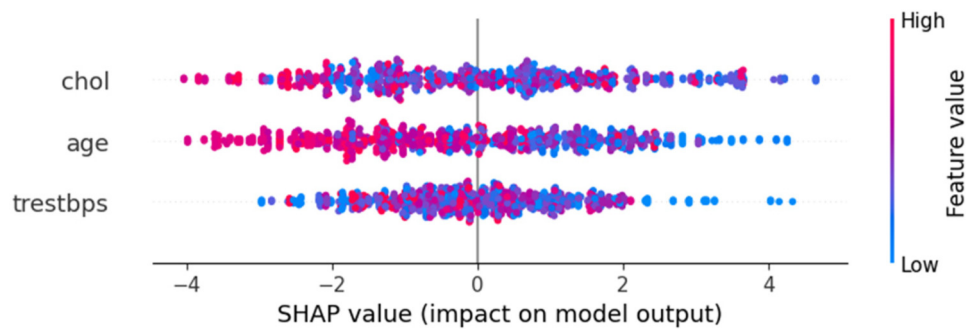


FIGURE 8 SHAP summary plot.

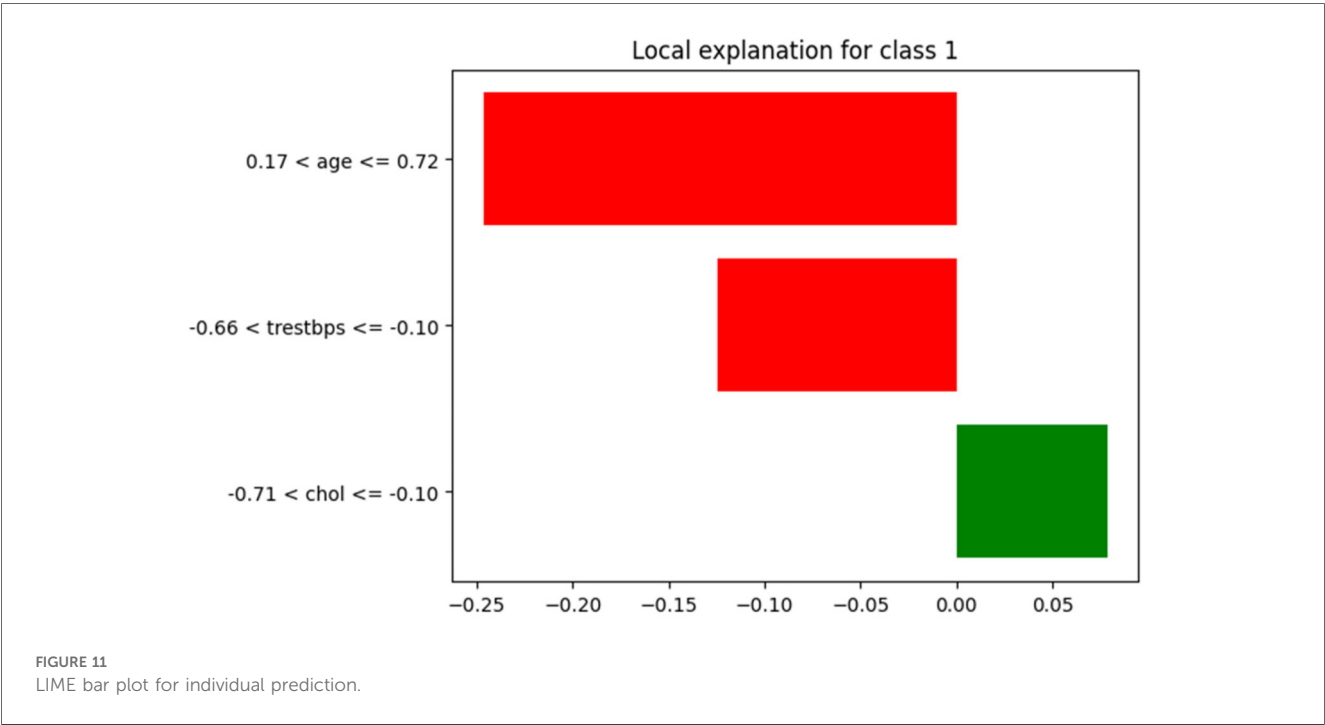
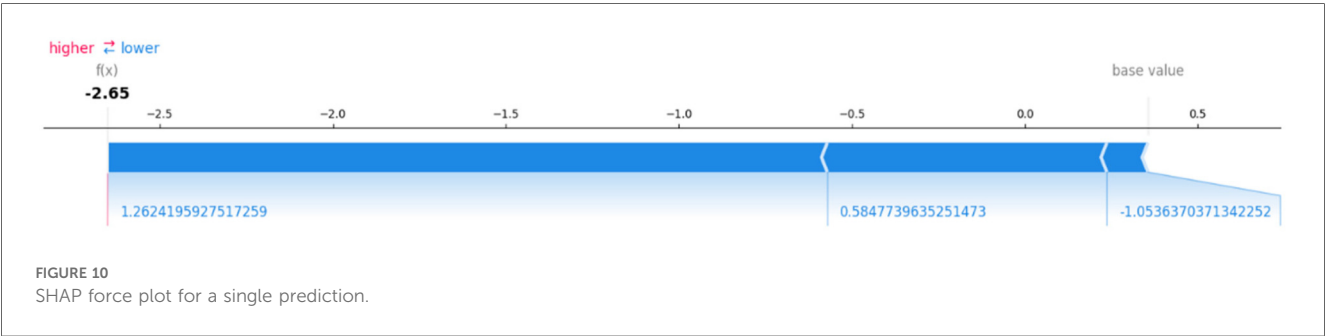
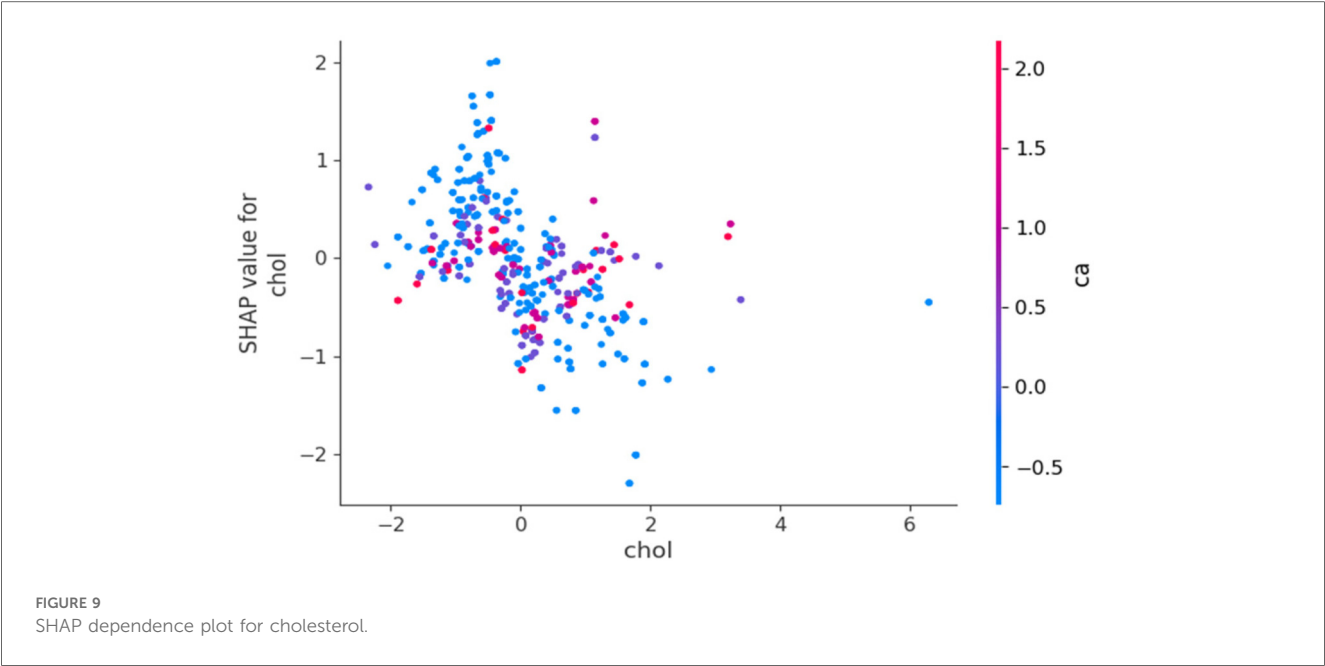


TABLE 6 Comparison of the proposed model with state-of-the-art models for HD prediction.

S. No.	Reference	Technique	Advantages	Limitations/future work	Accuracy (%)
1	Narayanan (32)	RF with SMOTE	High accuracy, sensitivity, and specificity in cardiac disease prediction using UCI heart dataset.	Focused only on the UCI dataset; more diverse datasets should be evaluated.	96.6
2	Ma et al. (20)	FlexiPulse sensor with ML	Accurate and affordable wearable tech for HD monitoring; uses ML for diagnostics	Needs large-scale validation for real-world implementation.	98.7
3	Kanagarathinam et al. (35)	Hybrid dataset with CatBoost	Achieved high accuracy using CV on the “Sathvi” dataset	Limited exploration of alternative classifiers and datasets.	98.11
4	Faizal et al. (22)	XG for heart failure prediction	Promising results for early mortality estimation and public health enhancement	Limited accuracy; opportunities exist to improve generalizability and robustness.	86.36
5	Li et al. (33)	Fusion models for CVD severity prediction	High accuracy in binary and multiclass classification, leveraging outputs from multiple algorithms.	Requires scalability and validation with larger datasets.	93–95 (binary); 72–75 (multiclass)
6	Proposed Optimized RST-ML Model	Integrating RST for feature selection with diverse ML classifiers and MCDM techniques.	Efficient feature selection with RST and enhanced accuracy through MCDM-based ranking (TOPSIS).	Needs validation on broader datasets and real-world clinical scenarios.	98.70

Wisconsin Diagnostic dataset, and CKD dataset. These datasets were chosen due to their varying characteristics, allowing for a comprehensive assessment of the proposed method's robustness.

The Obesity Levels dataset (56) comprises 2,111 records collected from Mexico, Peru, and Colombia. It includes 17 attributes related to eating habits and physical activities, classifying individuals into different obesity levels. A significant portion (77%) of this dataset was synthetically generated using the SMOTE technique, while the remaining 23% was collected directly from users via a web platform. The dataset provides a real-world scenario with categorical, binary, and continuous features, making it a challenging yet valuable resource for obesity classification.

The Breast Cancer Wisconsin Diagnostic dataset (57) contains 569 instances and 30 real-valued features extracted from digitized images of fine needle aspirates of breast masses. These features describe the characteristics of cell nuclei and are used to classify tumors as benign or malignant. The dataset does not contain missing values, making it suitable for evaluating the classification accuracy of various ML models.

The CKD dataset (58) consists of 400 records and 25 features, including critical health indicators such as red blood cell count, white blood cell count, and blood pressure levels. The target variable, “classification,” indicates whether a patient has CKD or not. The dataset required preprocessing due to missing values, and all rows containing NaNs were removed as per standard data cleaning protocols. This dataset is particularly relevant to the study due to its medical nature and direct implications for HD prediction. The performance of the proposed method was validated using these datasets by evaluating key performance metrics. The results are presented in the Table 7.

The proposed approach achieved outstanding performance. The Obesity Levels dataset, the model attained 95.51% accuracy, demonstrating its effectiveness in handling synthetic and real-world data. The Breast Cancer Wisconsin dataset yielded 98.33% accuracy, indicating the model's ability to distinguish between benign and malignant tumors. Finally, for the CKD dataset, the model achieved an accuracy of 96.49%, proving its capability in medical diagnostics. These results highlight the effectiveness of the proposed Optimized

TABLE 7 Performance comparison of the proposed method on different datasets.

Parameters	Obesity levels	Breast cancer Wisconsin	Chronic kidney disease
Ac(%)	95.51	98.33	96.49
Pr	0.94	0.98	0.95
Sn	0.95	0.99	0.95
F1	0.94	0.98	0.95
MCC	0.93	0.97	0.92
MAE	0.09	0.03	0.03
Tt (s)	0.23	0.06	0.12

RST-ML approach in various domains, particularly in medical and health-related datasets, showcasing its potential for real-world applications in disease prediction and classification.

4 Discussion

The proposed work demonstrates the potential of a stacked ensemble model combined with RST for accurate and interpretable prediction of HD risk. Several important methodological choices and limitations are discussed below, alongside directions for future research.

The adoption of RST for feature selection over modern feature selection techniques due to its inherent ability to handle uncertainty and imprecision without requiring prior data assumptions or parameter tuning. Unlike statistical or embedded methods such as Lasso, RST identifies minimal attribute subsets (reducts) based on data dependencies, thereby preserving classification accuracy while offering high interpretability and transparency which is critical in healthcare applications. Compared to statistical or black-box techniques like Lasso or embedded models, RST provides a rule-based, explainable framework while preserving classification accuracy during dimensionality reduction. This makes RST, and Johnson's Algorithm in particular, a robust and suitable choice for the proposed Optimized RST-ML model.

To ensure the robustness of the ensemble learning framework, each base model in the best-performing stack, Stack 4, was independently optimized using GridSearchCV. However, due to computational constraints and the complexity of hierarchical model tuning, hyperparameter optimization was not extended to the other stacked ensembles. Fine-tuning the final stacked models could potentially enhance predictive performance and may also influence the outcomes of the MCDM-TOPSIS ranking procedure used for model selection. Future work will investigate this possibility to assess whether ensemble-level optimization yields improved or different performance rankings.

Although the current analysis is limited to clinical datasets, the growing importance of non-invasive and wearable sensor data in cardiovascular health monitoring is recognized. Real-time data streams from such technologies could significantly enrich early detection and personalized risk stratification. Future research will incorporate these emerging data sources to enhance the clinical relevance and dynamic applicability of the proposed framework.

Furthermore, Country-level validation could offer deeper insights into model generalisability, as differences in population health indicators, healthcare systems, and risk factor distributions across countries may affect performance. Due to the limited availability of complete records, such validation was not feasible in the proposed work. Future work will consider larger, more balanced international datasets to assess cross-regional robustness and identify potential biases in predictive accuracy.

In summary, this study lays the foundation for a scalable and explainable HD prediction model. By integrating RST-based feature selection, ensemble learning, and multi-criteria decision-making, the study provides a comprehensive and transparent approach that balances accuracy, interpretability, and clinical relevance. Future research will focus on enhancing generalisability, leveraging wearable health data, and refining ensemble tuning to further strengthen the utility of the proposed model in diverse healthcare settings.

5 Conclusion and future work

This study proposed a hybrid framework that integrates RST for feature selection, ensemble ML through stacking, and MCDM using the TOPSIS technique to enhance classification performance in HD prediction. By employing Johnson's algorithm, the model effectively reduced dimensionality by preserving critical features. Multiple ML classifiers were trained and evaluated, and their strengths were assessed through five stacking configurations (Stack-1 to Stack-5). Pearson correlation analysis was used to ensure synergy among base learners in each stack, minimizing redundancy and maximizing learning diversity.

To identify the best-performing ensemble, a structured decision-making strategy was employed using the TOPSIS method, supported by MEREC-based weighting for objective criteria evaluation. Among the evaluated stacks, Stack-4 comprising advanced tree-based classifiers achieved the highest performance. Stack-4 was fine-tuned using GridSearchCV, with XG emerging as the best-performing model, achieving an

accuracy of 98.70% and 100% precision, thereby demonstrating its effectiveness in high-stakes medical diagnostic tasks. To ensure interpretability, XAI techniques were utilized to analyse feature contributions and model behaviour, offering transparent insights crucial for clinical decision-making.

The framework was validated not only on the HD dataset but also across diverse domains using publicly available datasets such as Obesity Levels, Breast Cancer Wisconsin, and CKD. The consistently high performance across all datasets demonstrates the model's generalizability and adaptability. This highlights the strength of integrating RST, ML, and MCDM into a single unified diagnostic framework, which is relatively unexplored in current medical AI applications.

Future research will focus on validating the framework using large-scale, real-world clinical data to assess its robustness and scalability. Furthermore, integrating non-invasive and wearable sensor data holds potential to enhance personalized monitoring and early detection. Lastly, although classical ML algorithms were used in this study, future work can explore DL integration with RST to improve feature learning and prediction performance in complex, high-dimensional datasets. Together, these directions aim to strengthen the clinical applicability and generalisability of the proposed model across diverse healthcare settings.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

TA: Writing – original draft, Writing – review & editing. GH: Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- World Health Organization (WHO). *Health topics*. Geneva: World Health Organization (2022). Available at: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 (Accessed June 14, 2025).
- Zhou X, Cai F, Li S, Li G, Zhang C, Xie J, et al. Machine learning techniques for prediction in pregnancy complicated by autoimmune rheumatic diseases: applications and challenges. *Int Immunopharmacol*. (2024) 134:112238. doi: 10.1016/j.intimp.2024.112238
- Swathy M, Saruladha K. A comparative study of classification and prediction of cardiovascular diseases (CVD) using machine learning and deep learning techniques. *ICT Express*. (2022) 8(1):109–16. doi: 10.1016/j.icte.2021.08.021
- Li JP, Haq AU, Din SU, Khan J, Khan A, Saboor A. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*. (2020) 8:107562–82. doi: 10.1109/ACCESS.2020.3001149
- Badawy M, Ramadan N, Hefny HA. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *J Electric Syst Inform Technol*. (2023) 10(1):40. doi: 10.1186/s43067-023-00108-y
- Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. *Heart Dis*. (2015) 7(1):129–37. doi: 10.090592/IJCSC.2016.018
- Solanki A, Barot MP. Study of heart disease diagnosis by comparing various classification algorithms. *Int J Eng Adv Technol*. (2019) 8(2S2):40–2. <https://www.ijeat.org/wp-content/uploads/papers/v8i2s2/B10090182S219.pdf>
- Kavitha R, Kannan E. An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*. IEEE (2016). p. 1–5
- Amin MS, Chiam YK, Varathan KD. Identification of significant features and data mining techniques in predicting heart disease. *Telemat Inform*. (2019) 36:82–93. doi: 10.1016/j.tele.2018.11.007
- Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng*. (2014) 40(1):16–28. doi: 10.1016/j.compeleceng.2013.11.024
- Greco S, Matarazzo B, Slowinski R. Rough sets theory for multicriteria decision analysis. *Eur J Oper Res*. (2001) 129(1):1–47. doi: 10.1016/S0377-2217(00)00167-3
- Zolfani SH, Görçün ÖF, Kundu P, Kükükönder H. Container vessel selection for maritime shipping companies by using an extended version of the grey relation analysis (GRA) with the help of type-2 neutrosophic fuzzy sets (T2NFN). *Comput Indust Engin*. (2022) 171:108376. doi: 10.1016/j.cie.2022.108376
- Karavidić Z, Projović D. A multi-criteria decision-making (MCDM) model in the security forces operations based on rough sets. *Decis Making Appl Manag Engin*. (2018) 1(1):97–120. doi: 10.31181/dmame180197k
- Kundu J, Kundu S. Cardiovascular disease (CVD) and its associated risk factors among older adults in India: evidence from LASI wave 1. *Clin Epidemiol Glob Health*. (2022) 13:100937. doi: 10.1016/j.cegh.2021.100937
- Day JR, Gupta A, Abro C, Jung K, Krishnamurti L, Takemoto C, et al. Risk factors and cardiovascular disease (CVD) related outcomes in hospitalized patients with hemophilia 10-year follow-up. *Blood*. (2020) 136:30–1. doi: 10.1182/blood-2020-136486
- Najafi A, Nemati A, Ashrafzadeh M, Zolfani SH. Multiple-criteria decision making, feature selection, and deep learning: a golden triangle for heart disease identification. *Eng Appl Artif Intell*. (2023) 125:106662. doi: 10.1016/j.engappai.2023.106662
- Baviskar V, Verma M, Chatterjee P, Singal G. Efficient heart disease prediction using hybrid deep learning classification models. *IRBM*. (2023) 44(5):100786. doi: 10.1016/j.irbm.2023.100786
- Mardani A, Hooker RE, Ozkul S, Yifan S, Nilashi M, Sabzi HZ, et al. Application of decision making and fuzzy sets theory to evaluate the healthcare and medical problems: a review of three decades of research with recent developments. *Expert Syst Appl*. (2019) 137:202–31. doi: 10.1016/j.eswa.2019.07.002
- Khan B, Riaz Z, Khoo BL. Advancements in wearable sensors for cardiovascular disease detection for health monitoring. *Mater Sci Engin R Rep*. (2024) 159:100804. doi: 10.1016/j.mser.2024.100804
- Ma Z, Hua H, You C, Ma Z, Guo W, Yang X, et al. Flexipulse: a machine-learning-enabled flexible pulse sensor for cardiovascular disease diagnostics. *Cell Rep Phys Sci*. (2023) 4(12):101690. doi: 10.1016/j.xcrp.2023.101690
- Hassan MM, Zaman S, Rahman MM, Bairagi AK, El-Shafai W, Rathore RS, et al. Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. *Comput Electric Engin*. (2024) 115:109130. doi: 10.1016/j.compeleceng.2024.109130
- Faizal ASM, Thevarajah TM, Khor SM, Chang SW. A review of risk prediction models in cardiovascular disease: conventional approach vs. Artificial intelligent approach. *Comput Methods Programs Biomed*. (2021) 207:106190. doi: 10.1016/j.cmpb.2021.106190
- Abdar M, Książek W, Acharya UR, Tan RS, Makarenkov V, Plawiak P. A new machine learning technique for an accurate diagnosis of coronary artery disease. *Comput Methods Programs Biomed*. (2019) 179:104992. doi: 10.1016/j.cmpb.2019.104992
- Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. *Med*. (2021) 2(6):642–65. doi: 10.1016/j.medj.2021.04.006
- Khalfan AF, Campisi SC, Lo RF, McCrindle B, Korczak DJ. Identifying cardiovascular disease risk endotypes of adolescent major depressive disorder using exploratory unsupervised machine learning. *JAACAP Open*. (2025) 3(2):291–301. doi: 10.1016/j.jaacop.2024.04.004
- Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Med Eng Phys*. (2022) 105:103825. doi: 10.1016/j.medengphy.2022.103825
- Venkat V, Abdelhalim H, DeGroat W, Zeeshan S, Ahmed Z. Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine. *Genomics*. (2023) 115(2):110584. doi: 10.1016/j.ygeno.2023.110584
- Wang H, Tucker WJ, Jonnagaddala J, Schutte AE, Jalaludin B, Liaw ST, et al. Using machine learning to predict cardiovascular risk using self-reported questionnaires: findings from the 45 and up study. *Int J Cardiol*. (2023) 386:149–56. doi: 10.1016/j.ijcard.2023.05.030
- Bhan A, Mangipudi P, Goyal A. An assessment of machine learning algorithms in diagnosing cardiovascular disease from right ventricle segmentation of cardiac magnetic resonance images. *Healthcare Analytics*. (2023) 3:100162. doi: 10.1016/j.health.2023.100162
- Qiu H, Yang P, Wang L. DCNet: a disease comorbidity network-based temporal deep learning framework to predict cardiovascular risk in patients with mental disorders. *Expert Syst Appl*. (2024) 254:124312. doi: 10.1016/j.eswa.2024.124312
- Li F, Wu P, Ong HH, Peterson JF, Wei WQ, Zhao J. Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. *J Biomed Inform*. (2023) 138:104294. doi: 10.1016/j.jbi.2023.104294
- Narayananana J. Implementation of efficient machine learning techniques for prediction of cardiac disease using SMOTE. *Procedia Comput Sci*. (2024) 233:558–69. doi: 10.1016/j.procs.2024.03.245
- Li P, Hu Y, Liu Z-P. Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods. *Biomed Signal Process Control*. (2021) 66:102474. doi: 10.1016/j.bspc.2021.102474
- Kibria HB, Matin A. The severity prediction of the binary and multi-class cardiovascular disease—A machine learning-based fusion approach. *Comput Biol Chem*. (2022) 98:107672. doi: 10.1016/j.compbiolchem.2022.107672
- Kanagarathinam K, Sankaran D, Manikandan R. Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. *Data Knowl Eng*. (2022) 140:102042. doi: 10.1016/j.datak.2022.102042
- Pan C, Poddar A, Mukherjee R, Ray AK. Impact of categorical and numerical features in ensemble machine learning frameworks for heart disease prediction. *Biomed Signal Process Control*. (2022) 76:103666. doi: 10.1016/j.bspc.2022.103666
- Mariadoss S, Augustin F. Enhanced sugeno fuzzy inference system with fuzzy AHP and coefficient of variation to diagnose cardiovascular disease during pregnancy. *J King Saud University-Comput Inform Sci*. (2023) 35(8):101659. doi: 10.1016/j.jksuci.2023.101659
- Lapp D. Heart disease dataset. Kaggle. Available at: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset> (Accessed: June 11, 2025).
- Pawlak Z. Rough sets. *Int J Parallel Program*. (1982) 11(5):341–56. doi: 10.1007/BF01001956
- Johnson DS. pproximation algorithms for combinatorial problems. In: *Proceedings of the Fifth Annual ACM Symposium on Theory of Computing*; 1973 April 30–May 02; Austin, Texas, USA. ACM. (1973). p. 38–49. doi: 10.1145/800125.804034

41. Hastie T, Tibshirani R, Friedman J, Franklin J. The elements of statistical learning: data mining, inference and prediction. *Mathematic Intell.* (2005) 27(2):83–5. doi: 10.1007/BF02985802
42. James G, Witten D, Hastie T, Tibshirani R. *An introduction to Statistical Learning*. New York: Springer (2013). 112, 18.
43. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer (2009). 2, 1–758.
44. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat.* (1992) 46(3):175–85. doi: 10.1080/00031305.1992.10475879
45. Yue S, Li P, Hao P. SVM Classification: its contents and challenges. *Appl Math A J Chinese Universities.* (2003) 18:332–42. doi: 10.1007/s11766-003-0059-5
46. Rish I. An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop Empirical MethodsArtific Intell.* (2001) 3(22):41–6. Available at: <https://faculty.cc.gatech.edu/~isbell/reading/papers/Rish.pdf>
47. Mitchell T. Decision tree learning. In: *Machine Learning*. (1997) 414:52–78.
48. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
49. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). p. 785–94
50. Schapire RE. Explaining adaboost. In: Schölkopf B, Luo Z, Vovk V, editors. *Empirical Inference*. Berlin, Heidelberg: Springer (2013). p. 37–52. doi: 10.1007/978-3-642-41136-6_5
51. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* (2006) 63:3–42. doi: 10.1007/s10994-006-6226-1
52. Zhou ZH. *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: CRC Press (2012).
53. Keshavarz-Ghorabae M, Amiri M, Zavadskas EK, Turskis Z, Antucheviciene J. Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry (Basel).* (2021) 13(4):525. doi: 10.3390/sym13040525
54. Chakraborty S. TOPSIS And modified TOPSIS: a comparative analysis. *Decis Analytics J.* (2022) 2:100021. doi: 10.1016/j.dajour.2021.100021
55. Ahmad GN, Fatima H, Ullah S, Saidi A. S.: efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *IEEE Access.* (2022) 10:80151–73. doi: 10.1109/ACCESS.2022.3165792
56. Mehrparvar F. Obesity levels. Kaggle. Available at: <https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels> (Accessed June 11, 2025).
57. Wolberg W, Mangasarian O, Street N, Street W. Breast Cancer Wisconsin (Diagnostic). (1993). Available at: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic> (Accessed June 11, 2025).
58. Iqbal M. Chronic kidney disease dataset. Kaggle. Available at: <https://www.kaggle.com/datasets/mansoordaku/ckdisease> (Accessed June 11, 2025).