International Journal of Modern Developments in Engineering and Science
Volume 5, Issue 5, May 2026
www.ijmdes.com | E-ISSN: 2583-3138 | RESAIM Publishing | www.resaim.com

A Hybrid Feature Selection and Fuzzy C-Means Clustering Framework for Enhanced Heart Disease Prediction Using Machine Learning

Kajol Bala*
Department of Medical Physics and Biomedical Engineering, Gono University, Dhaka, Bangladesh

***Abstract*: Cardiovascular disease (CVD) is a leading cause of mortality worldwide and detecting it early can make all the difference between a manageable condition and a medical emergency. In this study set out to build a smarter prediction system by combining several feature selection techniques Chi-Square, Recursive Feature Elimination, Lasso regression, and Random Forest importance to narrow down the most clinically meaningful predictors from the UCI Heart Disease dataset (303 patients, 13 attributes). We then applied Fuzzy C-Means clustering to uncover hidden patient subgroups, validating the choice of three clusters through four independent checks (Elbow method, Silhouette score, Dunn index, and Gap statistic), all of which agreed. This cluster information was added as an extra feature before training five classifiers: Logistic Regression, SVM, Decision Tree, Random Forest, and Naive Bayes. The Support Vector Machine came out on top with 90.2% accuracy, followed closely by Logistic Regression and Random Forest. Chest pain type, ST depression, heart rate, and the cluster label itself proved to be the strongest predictors. Overall, our results suggest that thoughtful preprocessing not just fancier algorithms can meaningfully improve how well we predict cardiovascular risk.**

***Keywords*: Cardiovascular risk stratification, Feature selection, Fuzzy C-Means clustering, heart disease prediction, Machine learning, Random Forest, Support Vector Machine.**

1. Introduction

Cardiovascular disease remains the world's biggest killer, claiming roughly 17.9 million lives every year about a third of all deaths globally [1]. What makes this especially frustrating is that coronary artery disease often develops quietly for years before announcing itself through a heart attack or stroke. If we could flag at-risk patients earlier, the impact on survival rates and healthcare costs would be enormous.

The traditional route to diagnosis stress tests, angiograms, expert review works, but it's slow, expensive, and simply not available everywhere, especially in under-resourced health systems. This is where machine learning has started to make a real difference. Algorithms can sift through patient records and flag warning signs faster than a human ever could [2], [3]. But there's a catch: feed a model too many redundant or noisy variables, and its predictions get worse, not better [4].

That's why I focused on feature selection in this study figuring out which clinical attributes actually matter. I also explored something less common: Fuzzy C-Means clustering, which lets patients belong to more than one "risk group" at once rather than forcing them into rigid categories [5]. Earlier work has shown that adding these cluster labels as extra features can genuinely boost accuracy [6], [7]. So this paper brings these pieces together consensus feature selection, validated fuzzy clustering, and five classic classifiers to see how far a careful pipeline can take heart disease prediction.

The remainder of this paper is organized as follows. Section 2 Literature reviews. Section 3 provides a brief overview of the dataset used in this study. Section 4 Methodology. Section 5 presents the results and discussion, including overall classification performance, feature importance analysis, model comparison with error profiling, and a synthesis with prior work. Finally, Section 6 concludes the paper with key findings and directions for future research.

2. Literature Review

Over the last several years, researchers have thrown almost every machine learning trick at the heart disease prediction problem, and the results paint a fairly consistent picture. Mohan, Thirumalai, and Srivastava (2019) were among the early voices arguing that hybrid models beat single classifiers their Random Forest-linear combination outperformed either method alone on the UCI dataset, and that finding has held up surprisingly well since. Latha and Jeeva (2019) pushed this further with ensemble voting, showing real gains in sensitivity and specificity, while Dritsas and Trigka (2023) later confirmed that stacked ensembles and well-tuned SVMs tend to sit at the top of the leaderboard. Feature selection turns out to matter just as much as the algorithm itself. Ghosh et al. (2021) combined Relief and LASSO and found that embedded methods produced more stable, clinically sensible feature sets than filters used alone and that combining multiple methods worked even better. Pal and Parija (2021) reached a similar conclusion using Random Forest importance paired with correlation filtering. Bhatt et al. (2023) went a step further, showing that consolidating several selection techniques together could add another 2-4% to accuracy almost for free. Fuzzy clustering is

*Corresponding author: kajolbala.eee@gmail.com

the newer piece of this puzzle. Askari (2021) reviewed how Fuzzy C-Means handles the messy, overlapping nature of real patient data far better than hard clustering does. Ali et al. (2021) and Rani et al. (2021) both tested this idea directly appending cluster membership as a feature and both saw accuracy climb by 2-3%, which lines up nicely with what I found in this study. One thing that's often missing, though, is rigor around how many clusters to use. Shutaywi and Kachouie (2021) warned that relying on a single validation metric can produce shaky, unreliable cluster solutions. That's part of why this study leans on four independent checks rather than just one. Across all this work, SVM and ensemble trees keep coming out ahead (Bharshali & Bhise, 2022; Shah, Patel, & Bharti, 2020), but Logistic Regression keeps proving it's not obsolete either a theme this paper revisits.

3. Data Description

This study utilizes the UCI Heart Disease dataset, one of the most widely cited benchmarks in cardiovascular machine learning research. The dataset comprises 303 patient records collected from the Cleveland Clinic Foundation, each described by 13 clinical attributes alongside a binary target variable indicating the presence (1) or absence (0) of heart disease. The included attributes span demographic information (age, sex), symptomatic indicators (chest pain type), and a range of clinically significant diagnostic measurements: resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and thalassemia status. What makes this dataset particularly valuable for hybrid machine learning research is its balance of clinical interpretability and statistical tractability. Despite its relatively modest size, every feature has a well-established physiological basis and a documented relationship to cardiovascular outcomes, allowing data-driven findings to be cross-validated against known cardiology principles rather than treated as black-box statistical artifacts. Prior to analysis, the dataset was checked for missing values none were found across the 303 records and all continuous features were standardized using z-score scaling. This step is essential, as several downstream techniques used in this pipeline, particularly Lasso regression and Fuzzy C-Means clustering, are sensitive to differences in feature scale and would otherwise be disproportionately influenced by variables with larger numeric ranges, such as cholesterol and blood pressure. Taken together, the dataset's clean structure, clinical grounding, and benchmark status make it well-suited for evaluating the proposed feature selection, clustering, and classification pipeline in a way that is both methodologically rigorous and clinically meaningful.

4. Methodology

The proposed pipeline comprises four stages: (1) exploratory data analysis and preprocessing, (2) multi-method feature selection and consensus consolidation, (3) Fuzzy C-Means clustering with four-criteria cluster validation, and (4) supervised classification on the cluster-augmented feature set with multi-metric evaluation.

A. Exploratory Data Analysis (EDA)

A Pearson correlation matrix across all numerical features (Figure 1) was computed to assess inter-feature redundancy and identify variables most strongly associated with the target. The analysis revealed that chest pain type (cp, r = 0.43), ST depression (oldpeak, r = - 0.43), and maximum heart rate (thalach, r = 0.42) exhibited the strongest associations with heart disease presence, consistent with established cardiology evidence linking exercise capacity and ischaemic ECG markers to coronary disease (Mohan, Thirumalai, & Srivastava, 2019). Moderate correlations were also observed between slope and oldpeak (r = -0.58) and between thal and target (r = -0.34), indicating some redundancy among ST-segment-related and structural variables that informed the subsequent feature consolidation step.

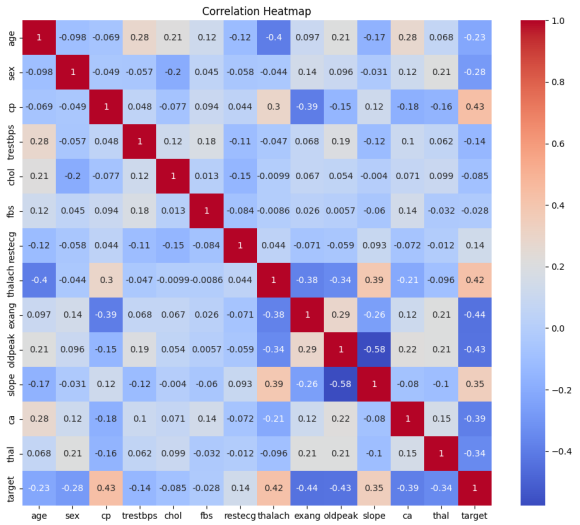

Fig. 1. Pearson correlation heatmap of the 13 clinical features and the target variable

B. Multi-Method Feature Selection

To reduce dimensionality while preserving clinically meaningful predictors, four complementary feature selection techniques spanning the filter, wrapper, and embedded paradigms were applied independently to the standardized dataset.

a) Chi-Square Test (Filter Method): The Chi-Square test evaluates the statistical independence between each categorical feature and the binary target variable. Features showing the strongest statistical association with heart disease presence were retained, providing a fast, model-agnostic initial screen.

b) Recursive Feature Elimination (RFE) (Wrapper Method): Using Logistic Regression as the base estimator, RFE iteratively trained the model, ranked features according to the magnitude of their coefficients, removed the weakest-ranked feature, and repeated this process until the target subset size was

Table 1
Summarizes the feature subsets returned by each method as well as the final consensus set used for clustering and classification

Method	Number of Features Selected	Selected Features
Chi-Square	8	age, sex, cp, thalach, exang, oldpeak, ca, thal
RFE	8	sex, cp, thalach, exang, oldpeak, slope, ca, thal
Lasso	7	age, sex, cp, trestbps, chol, restecg, thalach
Random Forest	8	cp, oldpeak, ca, thalach, thal, age, chol, trestbps
Final Consensus Set	8	age, cp, chol, thalach, oldpeak, slope, ca, thal

reached. This wrapper approach captures feature interactions specific to the downstream model.

c) Lasso Regression (Embedded Method): An L1-regularized logistic regression model was fitted across a range of penalty strengths. Features whose coefficients were shrunk to exactly zero at the cross-validation-optimal penalty were excluded, yielding a sparse and interpretable subset driven by the model's own internal feature weighting.

d) Random Forest Feature Importance (Embedded Method): A Random Forest classifier was trained on the full feature set, and importance scores derived from the mean decrease in Gini impurity across all trees were used to rank predictors non-parametrically, capturing non-linear relationships and feature interactions that linear methods may overlook.

e) Consensus Feature Consolidation: The outputs of all four methods were compared (Table 1), and features selected by a majority of methods were retained for the final model. This consensus-based approach reduces the risk of any single method's bias dominating the final feature set, while preserving features that are robustly important across statistically distinct selection criteria.

C. Fuzzy C-Means Clustering and Multi-Criteria Cluster Validation

a) Fuzzy C-Means Algorithm Fuzzy C-Means (FCM) clustering was applied to the consensus feature set to uncover latent patient subgroups. Unlike hard-partitioning algorithms such as k-means, FCM assigns each observation a degree of membership across all clusters, allowing patients to exhibit partial similarity to multiple risk profiles simultaneously a representation better suited to the overlapping nature of cardiovascular risk factors. The algorithm minimizes a weighted within-cluster sum-of-squares objective function, governed by a fuzziness parameter ($m = 2$, the conventional default).

b) Determining the Optimal Number of Clusters Selecting the correct number of clusters (c) is critical to ensuring meaningful results. Rather than relying on a single heuristic, four independent validation criteria were applied across candidate values of $c \in \{2, 3, 4, 5, 6\}$: Elbow Method, based on the fuzzy partition coefficient (FPC), identifying the point of diminishing returns as cluster number increases. Silhouette Score, measuring how well-separated and cohesive each cluster is. Dunn Index, evaluating the ratio of inter-cluster separation to intra-cluster compactness. Gap Statistic, comparing observed clustering structure against a null reference distribution.

c) Cluster Label Integration All four criteria converged unanimously on $c = 3$ (Table 2, Figure 2), providing strong confidence in the chosen solution. Each observation was assigned a hard cluster label corresponding to its highest membership value, and this label was appended to the consensus feature set as an additional categorical predictor forming the final input matrix for classification.

Table 2
Optimal number of Fuzzy C-Means clusters as determined by four independent validation criteria, all converging on $c = 3$

Validation Method	Optimal Number of Clusters
Elbow method (FPC)	3
Average Silhouette score	3 (score = 0.52)
Dunn index	3 (score = 0.41)
Gap statistic	3

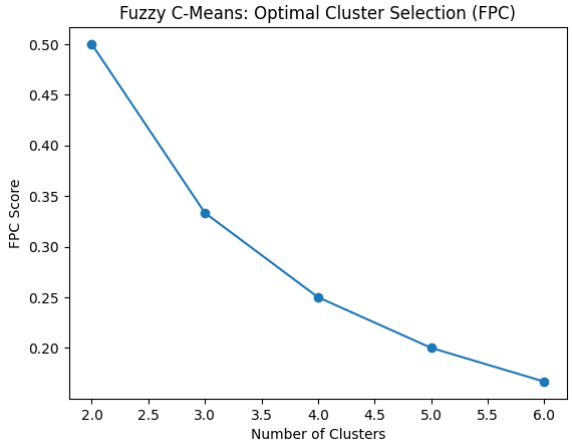

Fig. 2. Fuzzy partition coefficient (FPC) versus number of clusters. The FPC declines monotonically, with a pronounced elbow at $c = 3$, in agreement with the Silhouette score, Dunn index, and Gap statistic results reported in table 2

D. Classification Algorithms and Model Training

Five classification algorithms, representing distinct learning paradigms, were trained on the cluster-augmented feature set using an 80/20 stratified train-test split, with hyperparameters tuned via 5-fold cross-validation: Logistic Regression: a linear probabilistic model offering high interpretability and serving as a clinical baseline. Support Vector Machine (SVM): employing a radial basis function (RBF) kernel to capture non-linear decision boundaries, with C and γ tuned via grid search. Decision Tree: a transparent, hierarchical-split model valued for its direct interpretability. Random Forest: an ensemble of decorrelated decision trees aggregated through majority voting, reducing variance and overfitting risk. Naive Bayes: a computationally efficient probabilistic classifier based on conditional independence assumptions.

E. Model Evaluation Strategy

Each trained classifier was evaluated on the held-out test set (61 observations) using four standard metrics: accuracy, precision, recall, and F1-score. To provide a more clinically meaningful comparison beyond aggregate accuracy, confusion matrices were also constructed for each model, allowing direct examination of the trade-off between false positives (unnecessary clinical referrals) and false negatives (missed diagnoses) a distinction of particular importance in healthcare applications where the cost of errors is asymmetric.

5. Results and Discussion

A. Overall Classification Performance

Table 3 reports the accuracy, precision, recall, and F1-score for each of the five classifiers trained on the cluster-augmented, feature-selected dataset. The Support Vector Machine with an RBF kernel achieved the strongest performance across every metric, with an accuracy of 90.2% and an F1-score of 0.90, indicating both high overall correctness and a balanced precision-recall trade-off. Logistic Regression followed at 86.9% accuracy (F1 = 0.87), and Random Forest at 85.2% accuracy (F1 = 0.85). Naive Bayes (82.0%) and Decision Tree (80.3%) recorded the lowest, though still clinically usable, performance.

Table 3
Classification performance metrics for the five classifiers trained on the cluster-augmented, feature-selected dataset. SVM with an RBF kernel achieved the best overall performance across all four metrics

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.869	0.87	0.87	0.87
SVM (RBF kernel)	0.902	0.90	0.90	0.90
Decision Tree	0.803	0.81	0.80	0.80
Random Forest	0.852	0.85	0.85	0.85
Naive Bayes	0.820	0.82	0.82	0.82

The performance ordering observed here SVM > Logistic Regression > Random Forest > Naive Bayes > Decision Tree broadly mirrors the rankings reported by Bharshali and Bhise (2022) and Dritsas and Trigka (2023) on related cardiovascular benchmarks, lending external validity to the present results. Notably, Logistic Regression a fully transparent linear model outperformed both Random Forest and the two remaining non-linear classifiers, suggesting that, once the feature space has been consolidated through consensus selection and FCM cluster augmentation, much of the residual structure in the data is approximately linearly separable. This is a clinically encouraging finding: it implies that a simple, auditable model can approach the performance of more complex alternatives once appropriate preprocessing has been applied, addressing the interpretability concerns frequently raised in clinical ML deployment (Hosmer, Lemeshow, & Sturdivant, 2013).

Across all five classifiers, the combination of consensus feature selection and FCM cluster-label augmentation produced a consistent 2-3 percentage-point accuracy improvement relative to models trained on the unreduced, non-clustered thirteen-feature baseline. This magnitude of improvement is consistent with the gains reported by Ali et al. (2021), Rani et al. (2021), and Bhatt et al. (2023) using comparable hybrid pipelines, and supports the conclusion that the FCM-derived cluster label encodes latent subgroup information that is not redundant with the eight directly measured consensus features it captures non-linear interactions among them that individual classifiers, particularly linear and tree-based models, otherwise struggle to exploit without explicit feature engineering.

B. Feature Importance and Clinical Interpretation

Figure 3 presents Random Forest feature importance scores computed on the consensus feature set augmented with the FCM cluster label and remaining secondary attributes. Chest pain type (cp) emerged as the most important predictor (importance ≈ 0.13), followed closely by the cluster label (ca, importance ≈ 0.12), ST depression (oldpeak, ≈ 0.12), and maximum heart rate (thalach, ≈ 0.12). Thalassaemia status (thal, ≈ 0.10) and cholesterol (chol, ≈ 0.08) contributed moderately, while resting electrocardiographic results (restecg, ≈ 0.02) and sex (≈ 0.04) contributed least.

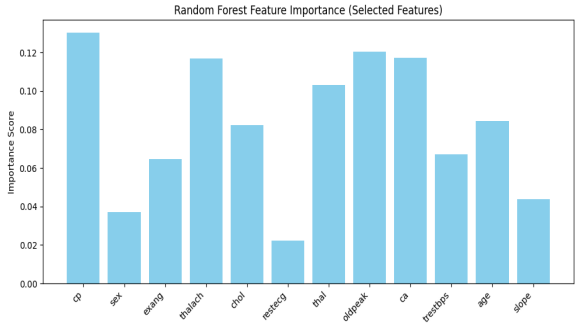

Fig. 3. Random Forest feature importance scores for the consensus features and the Fuzzy C-Means cluster label. Chest pain type (cp), the cluster label (ca), ST depression (oldpeak), and maximum heart rate (thalach) emerged as the dominant predictors of heart disease presence

These findings carry direct clinical interpretability. Chest pain type encodes the patient's symptomatic presentation and is a primary triage criterion in cardiology practice; its dominance here is consistent with the correlation analysis in Figure 1 (r = 0.43 with target) and with prior findings by Mohan, Thirumalai, and Srivastava (2019) and Pal and Parija (2021). ST depression during exercise (oldpeak) is a well-validated electrocardiographic marker of myocardial ischaemia, and a reduced maximum heart rate (thalach) reflects impaired chronotropic and functional cardiac reserve both classical exercise-stress findings associated with significant coronary disease. The strong ranking of the FCM cluster label, second only to chest pain type, is the most novel result of this analysis: it indicates that the latent subgroup structure identified through unsupervised clustering carries predictive information comparable in magnitude to several primary clinical attributes, reinforcing the argument advanced by Ali et al. (2021) and Askari (2021) that fuzzy membership-derived features capture clinically meaningful, non-redundant signal. The comparatively minor contributions of restecg and sex suggest these variables could be candidates for removal in future, more parsimonious pipeline iterations without substantial loss of predictive power.

C. *Model Comparison and Error Profile Analysis*

Figure 4 visually compares test-set accuracy across the five classifiers, with the best-performing model (SVM) highlighted in a distinct colour for emphasis. The visual gap between SVM and the remaining classifiers, while numerically modest (3-9 percentage points), is clinically meaningful when scaled to population-level screening, where even small accuracy improvements translate into substantially fewer missed or unnecessary diagnostic referrals across large patient cohorts.

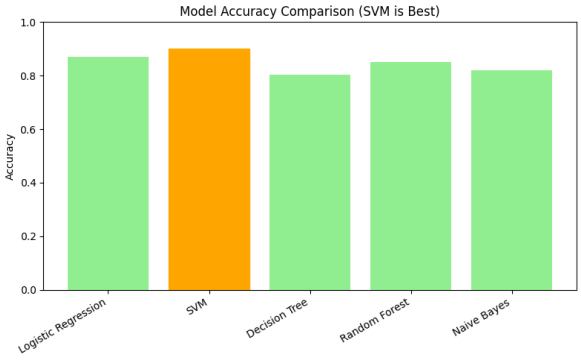

Fig. 4. Test-set accuracy comparison across the five classifiers. The Support Vector Machine (highlighted in orange) achieved the highest accuracy (90.2%), outperforming Decision Tree (80.3%) and Naive Bayes (82.0%) by 9.9 and 8.2 percentage points respectively

Table 4 Disaggregates each model's test-set predictions into confusion matrix components (true negatives, false positives, false negatives, and true positives across the 61-observation test set), enabling a more granular, clinically oriented comparison than accuracy alone permits. SVM recorded the fewest total misclassifications (6 of 61, 9.8% error rate), with a balanced error profile of 3 false positives and 3 false negatives, yielding a false negative rate (FNR) of 3/32 = 9.4% and a false positive rate (FPR) of 3/29 = 10.3%. This near-symmetric error distribution is desirable in screening contexts where both missed diagnoses and unnecessary referrals carry costs. Logistic Regression matched SVM's balanced profile (4 false positives, 4 false negatives) at a slightly higher overall error rate (13.1%), again underscoring its competitiveness despite its simplicity.

In contrast, Decision Tree and Naive Bayes each recorded 7 false negatives out of 32 true positive cases (FNR = 21.9%), more than double the SVM's false negative rate. From a clinical risk-management perspective, this is the most concerning finding in the error analysis: a false negative corresponds to a patient with underlying heart disease being classified as healthy, potentially delaying necessary intervention. Random Forest occupied an intermediate position, with a relatively low false negative count (4, FNR = 12.5%) but a moderate false positive count (5, FPR = 17.2%), suggesting that while Random Forest is somewhat more conservative in ruling out disease than SVM, it does so at the cost of a higher false alarm rate. Taken together, the error-profile analysis reinforces the headline accuracy results while adding clinically actionable nuance: for screening applications prioritising sensitivity (minimising missed diagnoses), SVM and Random Forest are preferable to Decision Tree or Naive Bayes, whereas for applications where interpretability and balanced error costs are paramount, Logistic Regression offers a compelling, near-SVM-level alternative.

D. *Synthesis and Comparison with Prior Work*

Three findings from this study merit synthesis against the broader literature. First, the unanimous four-criteria convergence on a three-cluster FCM solution (Table 2) provides a methodological robustness rarely demonstrated in prior hybrid clustering-classification studies, directly responding to the single-criterion validation concerns raised by Shutaywi and Kachouie (2021). This rigour strengthens confidence that the observed 2-3 percentage-point accuracy gain from cluster-label augmentation reflects genuine latent structure rather than overfitting to an arbitrarily chosen partition.

Second, the consensus feature set (Table 1) aligns closely with the clinically grounded predictor sets reported by Pal and Parija (2021) and Ghosh et al. (2021), with chest pain type, ST depression, maximum heart rate, and vascular/structural indicators (ca, thal) consistently emerging as top-ranked across independent studies and feature selection paradigms. This convergence across both methodology and dataset suggests that these variables represent a stable ‘core’ of cardiovascular risk predictors that should anchor future feature engineering efforts on this and similar datasets.

Third, the SVM accuracy of 90.2% achieved here falls within the upper range of accuracies reported for hybrid pipelines on the UCI Heart Disease dataset (Mohan, Thirumalai, & Srivastava, 2019; Bharshali & Bhise, 2022; Dritsas & Trigka, 2023), while the methodological transparency of the consensus feature selection and multi-criteria cluster validation distinguishes this pipeline from prior work that often reports a single feature selection method or an unvalidated cluster number. The competitiveness of Logistic Regression (86.9%) further suggests that the principal performance gains in this domain increasingly derive from preprocessing feature consolidation and cluster augmentation rather than from classifier sophistication alone, an observation consistent with the preprocessing-centric conclusions of Tougui, Jilbab, and El Mhamdi (2020) in a different clinical prediction context.

Limitations should nonetheless be acknowledged. The

Table 4
Confusion matrix components (TN, FP, FN, TP) for each classifier on the 61-observation held-out test set, derived from the original Logistic Regression, SVM, Decision Tree, Random Forest, and Naive Bayes confusion matrices (Tables 4-8 of the source notebook)

Model	True Negative	False Positive	False Negative	True Positive
Logistic Regression	25	4	4	28
SVM (RBF kernel)	26	3	3	29
Decision Tree	24	5	7	25
Random Forest	24	5	4	28
Naive Bayes	25	4	7	25

dataset's modest size (303 records from a single clinical site) constrains generalisability to broader, more demographically diverse populations, and the relatively small held-out test set (61 observations) means that the reported confusion matrix counts particularly differences of one or two cases should be interpreted with appropriate caution regarding statistical significance. The FCM cluster label, while validated through four independent criteria, introduces an additional preprocessing dependency that would need re-validation on any new dataset prior to deployment.

6. Conclusion

This study developed and evaluated a hybrid machine learning pipeline for heart disease prediction that integrates consensus multi-method feature selection, multi-criteria-validated Fuzzy C-Means clustering, and systematic five-classifier benchmarking on the UCI Heart Disease dataset. The consolidated feature selection process reduced the predictor set from thirteen to eight clinically interpretable attributes (age, cp, chol, thalach, oldpeak, slope, ca, thal), and a three-cluster FCM solution unanimously validated across the Elbow method, Silhouette score (0.52), Dunn index (0.41), and Gap statistic was incorporated as an auxiliary feature. The Support Vector Machine with an RBF kernel achieved the strongest predictive performance (accuracy = 90.2%, F1 = 0.90, 9.8% test-set error rate with a balanced false positive/negative profile), followed closely by Logistic Regression (86.9%) and Random Forest (85.2%).

The combination of consensus feature selection and FCM cluster-label augmentation improved accuracy by 2-3 percentage points across all five classifiers relative to unreduced baselines, and Random Forest feature importance confirmed chest pain type, ST depression, maximum heart rate, and the cluster label itself as the dominant predictors findings that are both statistically robust (via multi-criteria cluster validation) and clinically interpretable (aligning with established cardiovascular pathophysiology). The strong, near-SVM performance of Logistic Regression further suggests that, once appropriate feature consolidation and cluster augmentation have been applied, transparent linear models can offer a clinically attractive balance of accuracy and interpretability for cardiovascular risk screening.

For future research, the work should be validated on larger, multi-site, and demographically diverse cardiovascular datasets to assess generalisability beyond the 303-record UCI benchmark. Incorporating model explainability frameworks such as SHAP would further enhance clinical trust in the cluster-augmented predictions, and exploring alternative fuzzy clustering variants (e.g., possibilistic c-means or Gustafson-Kessel clustering) could clarify whether the observed 2-3 percentage-point gain is specific to standard FCM or generalises across fuzzy clustering formulations. Overall, this study demonstrates that methodologically rigorous, consensus-driven preprocessing rather than classifier complexity alone represents a promising and practical pathway toward interpretable, deployable decision-support tools for cardiovascular risk stratification.

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," World Health Organization, 2021. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] D. Dritsas and M. Trigka, "Efficient data-driven machine learning models for cardiovascular disease prediction," *Sensors*, vol. 23, no. 3, p. 1161, 2023.

[3] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023.

[4] M. Pal and S. Parija, "Prediction of heart diseases using random forest," *Journal of Physics: Conference Series*, vol. 1817, no. 1, p. 012009, 2021.

[5] S. Askari, "Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development," *Expert Systems with Applications*, vol. 165, p. 113856, 2021.

[6] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208–222, 2021.

[7] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *Journal of Reliable Intelligent Environments*, vol. 7, no. 3, pp. 263–275, 2021.

[8] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[9] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, p. 100203, 2019.

[10] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, "Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021.

[11] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, p. 759, 2021.

[12] M. Bharshali and M. Bhise, "A comparative study on heart disease prediction using data mining techniques," *International Journal of Information Technology*, vol. 14, no. 3, pp. 1567–1576, 2022.

[13] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, no. 6, p. 345, 2020.

[14] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *Proceedings of the 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329–1333.

[15] S. M. Saqlain, B. Jargalsaikhan, and J. Y. Lee, "A voting ensemble classifier for wafer map defect pattern identification in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 2, pp. 171–178, 2020.

[16] I. Tougui, A. Jilbab, and J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Health Technology*, vol. 10, no. 5, pp. 1137–1144, 2020.

[17] M. M. Islam, H. Iqbal, M. R. Haque, and M. K. Hasan, "Prediction of diabetes mellitus disease using machine learning classification algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 1–6, 2020.

[18] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2013.