(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 16, No. 3, 2025

Adaptive Ensemble Selection for Personalized Cardiovascular Disease Prediction Using Clustering and Feature Selection

Mutaz A. B. Al-Tarawneh[1], Khaled S. Al-Maaitah[2], Ashraf Alkhresheh[3]
College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait[1]
Computer Engineering Department, Mutah University, Karak, Jordan[2]
Computer Science Department, Tafila Technical University, Tafila, Jordan[3]

Abstract—**Cardiovascular disease (CVD) remains one of the leading causes of mortality worldwide, highlighting the need for early and precise prediction to support timely intervention. This study introduces an ensemble-based adaptive approach that personalizes CVD prediction by dynamically adjusting model configurations based on patient subgroups. To achieve this, various clustering techniques, including KMeans, DBSCAN, and MeanShift, are employed alongside feature selection methods such as chi-square, Mutual Information, and a baseline that incorporates all features. By tailoring classifier selection to each cluster, the proposed approach optimizes predictive performance, with ensemble models configured using Multi-Layer Perceptron (MLP) or Decision Tree classifiers. Through extensive experiments utilizing 10-fold cross-validation, results indicate that the adaptive ensemble consistently surpasses the static ensemble in key performance metrics, including accuracy, precision, recall, F1 score and AUC. In particular, the highest accuracy of 95.57% was achieved using MeanShift clustering with the entire set of features, demonstrating the effectiveness of density-based clustering in improving classification performance. Notably, this accuracy exceeds the best-reported results in previous studies, establishing a new benchmark for CVD prediction. These findings highlight the potential of adaptive ensemble selection to significantly improve diagnostic precision, providing valuable insights for personalized CVD prediction and broader applications in medical decision making.**

Keywords*—*Cardiovascular disease prediction; adaptive ensemble selection; clustering techniques; feature selection; personalized healthcare

I. INTRODUCTION

Cardiovascular disease (CVD) is a broad category of conditions that affect the heart and blood vessels, including hypertension, valvular disorders, arrhythmias, and coronary artery disease [1], [2]. As one of the leading causes of mortality worldwide, CVD underscores the urgent need for an early and accurate diagnosis to improve patient outcomes [3]. Traditional diagnostic approaches, such as analyzing vital signs, conducting physical examinations, and interpreting electrocardiograms, have proven effective but are often time consuming, prone to human error, and dependent on expert interpretation [4], [5]. These limitations can delay diagnosis and can lead to missed early indicators of disease progression. As a result, there is a growing demand for advanced diagnostic tools that can facilitate early detection and support timely clinical intervention [6].

Rapid advances in artificial intelligence (AI) and machine learning (ML) have opened new possibilities to automate and improve CVD diagnosis [7], [8]. ML algorithms excel at analyzing complex patterns in large-scale cardiac datasets, allowing more precise and data-driven predictions that aid clinical decision making [9]. A wide range of ML techniques, including logistic regression, k-nearest neighbors, decision trees, support vector machines, and ensemble models, have been successfully applied to CVD prediction [10], [11]. Among these, ensemble learning has gained significant traction due to its ability to combine multiple models, improving predictive accuracy and robustness for complex medical conditions like CVD [12], [13].

Despite these advancements, traditional ensemble models often rely on a fixed feature set, which may include irrelevant or redundant variables. This can lead to overfitting, reduced generalization, and increased computational complexity. Feature selection plays a crucial role in mitigating these challenges by identifying the most informative predictors, thereby enhancing model efficiency and improving diagnostic performance [14], [15], [16].

This study introduces an ensemble-based adaptive approach for CVD diagnosis that tailors model configurations to distinct subgroups of patients. By incorporating clustering techniques, patients are segmented into groups with shared characteristics, allowing the optimization of ensemble configurations based on the specific characteristics of each cluster. In addition, multiple feature selection techniques are applied and analyzed, including chi-square and mutual information, to assess their impact on predictive accuracy, alongside a baseline scenario where all features are retained. This comprehensive evaluation aims to highlight the role of feature selection in improving diagnostic reliability and efficiency.

The key contributions of this study include:

- The development of a dynamic ensemble-based CVD detection framework that adapts model selection based on patient clustering to enhance diagnostic performance.
- A comparative analysis of feature selection methods, examining their impact on model accuracy and efficiency in contrast to a baseline approach using all available features.

- A thorough evaluation of various clustering and ensemble configurations across multiple performance metrics to identify the most effective strategies for CVD diagnosis.

The remainder of this paper is organized as follows. Section II provides an overview of related work on ensemble-based ML models for CVD prediction. Section III details the methodology, covering data pre-processing, clustering, feature selection, and model training. Section IV presents the experimental results, followed by a discussion of key findings. Finally, Section V concludes with insight and implications for future research.

II. LITERATURE REVIEW

The application of machine learning (ML) in the diagnosis of cardiovascular disease (CVD) has gained significant attention in recent years due to its potential to enhance both accuracy and efficiency. Traditional diagnostic methods often depend on extensive clinical expertise and are susceptible to human error. To overcome these challenges, ML models have been increasingly employed to analyze complex clinical data, providing more reliable and data-driven predictions [17]. Among these approaches, ensemble learning has emerged as a powerful technique for integrating multiple base models, improving both prediction accuracy and robustness in CVD detection.

Ensemble learning combines predictions from multiple classifiers to enhance overall model performance, as demonstrated in recent studies exploring various voting and stacking strategies. For instance, the authors of [18] implemented a voting ensemble that integrated deep learning (DL) classifiers with traditional ML models, achieving an accuracy of 88.7% in heart disease prediction. Their approach used six classifiers: Random Forest (RF), k-Nearest Neighbors (KNN), Decision Tree (DT), Extreme Gradient Boosting (XGB), Deep Neural Network (DNN), and Kernel Deep Neural Network (KDNN). Similarly, studies in [10], [19] explored voting ensembles combining classifiers such as Naïve Bayes (NB), Artificial Neural Network (ANN), Logistic Regression (LR), DT, and KNN. These studies also incorporated extra tree feature selection, demonstrating improved accuracy on the Cleveland dataset.

The integration of feature selection with ensemble models has become a key research area, improving both model interpretability and computational efficiency. Selecting only the most relevant features reduces overfitting and enhances predictive performance. For example, [20] applied Chi-square and recursive feature elimination (RFE) together with ensemble methods, reporting that Classification and Regression Trees (CART) achieved the highest accuracy (87. 65%) in CVD prediction. Furthermore, [21] investigated the effects of combining bagging, boosting, majority voting, and stacking with feature selection in various base classifiers, including NB, RF, C4.5, Bayesian Network, Multilayer Perceptron (MLP), and Projective Adaptive Resonance Theory (PART), achieving an accuracy improvement of 7. 26% for weaker classifiers.

To further refine predictive accuracy, advanced optimization techniques have been integrated into the ensemble frameworks. For example, [22] explored the combination of correlation-based feature selection (CFS) with Particle Swarm Optimization (PSO), achieving an accuracy of 85.71% for CVD diagnosis. Similarly, [23] developed a voting ensemble incorporating Support Vector Machine (SVM), DT, and ANN classifiers, significantly outperforming individual models in precision, recall, and F1 score. Another study, [24], proposed a novel voting strategy using an ensemble of six ML models, achieving an accuracy of 83%, exceeding the performance of any single model.

Recent efforts have also incorporated deep learning techniques into ensemble frameworks to capture complex patterns in high-dimensional medical data. In [25], the authors combined Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks with traditional ML models such as RF, SVM and KNN in a voting ensemble, leading to an increase in accuracy of 2.1% compared to individual models in the Cleveland dataset. Similarly, [26] proposed an ensemble approach using SVM, NB and ANN classifiers with majority vote, reporting an accuracy of 87 05%.

One of the most recent advances in this field is presented in [27], where a voting ensemble approach was integrated with the selection of Chi-square characteristics for improved CVD detection. This model employed classifiers such as NB, RF, LR, and KNN, achieving an accuracy of 92.11% demonstrating the impact of feature selection in reducing computational overhead while improving predictive performance.

Although these studies highlight the effectiveness of ensemble learning for CVD prediction, most rely on static ensemble configurations that do not adapt to individual patient profiles. In addition, limited research has comprehensively examined the influence of different feature selection techniques, particularly in scenarios where all features are retained, within ensemble frameworks for CVD prediction. This study addresses these gaps by introducing a dynamic ensemble-based approach, where patient clustering is employed to segment individuals into subgroups, each with an optimized ensemble configuration. Furthermore, multiple feature selection techniques are assessed, offering a comparative analysis of their impact on predictive performance in CVD diagnosis.

III. METHODOLOGY

The pseudocode presented in Algorithm 1 outlines a systematic approach to evaluate machine learning models and ensemble configurations in the detection of cardiovascular disease. Each step is designed to build on the previous one, ensuring a comprehensive and robust model evaluation process. The methodology begins with data loading, preprocessing, and scaling to standardize the data set, ensuring consistency between models and minimizing bias caused by the varying range of features. Feature selection techniques are then applied to identify the most relevant predictors, reducing dimensionality, and improving computational efficiency. This step improves the effectiveness of both clustering and classification by focusing on the most informative features.

Once the feature selection process is complete, clustering techniques are used to segment patients into distinct groups. These clusters serve as the foundation for adaptive ensemble models, allowing classifier configurations to be optimized for each subgroup based on their unique characteristics. Following clustering, individual classifiers are trained, fine-tuned, and integrated into static and adaptive ensemble frameworks to improve predictive performance. To ensure the reliability and generalizability of the models, cross-validation is conducted across multiple performance metrics. Finally, the performance results are aggregated and stored for further analysis, enabling a comparative evaluation of different ensemble configurations and providing insights into their effectiveness in CVD prediction.

Algorithm 1 Main Steps of the Research Methodology

```
Set random seed for reproducibility
Load and Prepare the Dataset
Load data from CSV
Split data into training and test sets
Normalize feature values
Define Feature Selection, Clustering Methods, and Ensemble Selectors
Set Up Cross-Validation and Hyperparameter Grids for Base Models
for each feature selection method do
    Apply feature selection
    Train and tune individual classifiers
    Define static and adaptive ensemble configurations
    for each clustering method and ensemble selector do
        Apply clustering to training data
        Train ensemble selector based on clusters
        for each validation fold do
            Predict using dynamic adaptive ensemble
            Evaluate performance metrics
        end for
        Store results for dynamic ensemble
    end for
    Evaluate Static Ensemble with Cross-Validation
    for each validation fold do
        Predict using static ensemble
        Evaluate performance metrics
    end for
    Store results for static ensemble
end for
Save All Results to CSV
```

A. Data Collection

In this study, the Cleveland Heart Disease dataset, a widely used public dataset from the University of California at Irvine (UCI) Machine Learning Repository, was used to predict the probability of heart disease [28]. The data set comprises 303 patient records and 76 attributes, although most research efforts usually focus on a subset of 14 key features. These include 13 input variables: age, sex, cholesterol level, heart rate, type of chest pain, fasting blood sugar, blood pressure, resting ECG, exercise-induced angina, ST slope, ST depression, the number of vessels detected by fluoroscopy, and thalassemia status. The final attribute serves as the output variable, indicating the presence or absence of heart disease as a binary classification (0 or 1) [29]. A detailed description of these attributes is provided in Table I.

B. Data Preparation and Preprocessing

The first step in the methodology involves data set preparation and pre-processing, which serves as a crucial foundation for building robust and accurate machine learning models. The data set is first loaded using `pandas`, with the input features (X) and the target variable (y) carefully separated to ensure a clear distinction between predictive factors and disease classification. To standardize feature ranges and improve model performance, `MinMaxScaler` is applied, normalizing all feature values between 0 and 1. This scaling process not only facilitates faster model convergence, but also ensures that features contribute fairly to the learning process, ultimately enhancing the predictive accuracy of cardiovascular disease detection.

TABLE I. ATTRIBUTE INFORMATION FOR THE CLEVELAND HEART DISEASE DATASET

Attribute	Type	Details
Age	Num	Age (years)
Sex	Categorical	1: Male, 0: Female
Cp	Categorical	Chest pain type (4: asymptomatic, 3: non-anginal, 2: atypical, 1: typical)
Trestbps	Num	Resting BP (mmHg)
Chol	Num	Serum cholesterol (mg/dL)
Fbs	Categorical	Fasting blood sugar $>$ 120 mg/dL (1: true, 0: false)
Restecg	Categorical	ECG (2: LV hypertrophy, 1: ST-T abnormality, 0: normal)
Thalach	Num	Max heart rate
Exang	Categorical	Exercise-induced angina (1: yes, 0: no)
Oldpeak	Num	ST depression during exercise
Slope	Categorical	ST segment slope (3: downward, 2: flat, 1: upward)
Ca	Categorical	Major vessels (0-3) visualized by fluoroscopy
Thal	Categorical	Thallium test (7: reversible defect, 6: fixed defect, 3: normal)
Num	Categorical	Heart disease diagnosis (1: $>$ 50% narrowing, 0: $\leq$ 50%)

C. Feature Selection

Feature selection plays a crucial role in refining the input variables to include only the most relevant features, thus reducing dimensionality, minimizing noise, and improving computational efficiency. Three selection methods are examined: no feature selection, Chi-Squared [30], and Mutual Information [31]. The Chi-Squared method evaluates the independence between features and the target variable, selecting features that are most correlated with disease presence. Mutual information, alternatively, calculates the information shared between each feature and the target, highlighting the features with the highest contribution to accurate predictions. By identifying the optimal subset of features, this stage improves model focus and predictive power in cardiovascular disease detection.

D. Clustering Methods

In the context of adaptive learning, clustering provides an unsupervised approach to grouping data points based on inherent similarities, enabling the identification of underlying patterns in the data set. This study explores eight clustering methods-KMeans, Gaussian mixture model (GMM), DBSCAN, aggregative clustering, spectral clustering, meanshift, affinity propagation and fuzzy C-means. Each technique offers a distinct approach to data segmentation, capturing various clustering structures that may correspond to different risk profiles or disease stages.

- K-Means [32]: A widely used centroid-based clustering method that partitions data into k clusters by minimizing intra-cluster variance. The objective function is given by:

$$\sum_{j=1}^{k}\sum_{x_i\in c_j}\|x_i-\mu_j\|^2$$

where each data point x_i is assigned to the nearest cluster centroid μ_j. As a “hard” clustering method, K-Means is efficient for large datasets but assumes spherical clusters, which may limit performance on complex data distributions.

- Gaussian Mixture Model (GMM) [33]: A probabilistic clustering approach that models data as a mixture of multiple Gaussian distributions. Each data point is assigned a probability of belonging to each cluster, enabling “soft” assignments. The probability distribution is given by:

  $$P(x_i)=\sum_{j=1}^{k}\pi_j\mathcal{N}(x_i|\mu_j,\Sigma_j)$$

  where π_j is the weight of cluster j, and $\mathcal{N}(x_i|\mu_j,\Sigma_j)$ represents the Gaussian distribution with mean μ_j and covariance matrix Σ_j. GMM is effective for modeling elliptical clusters and capturing overlapping distributions.

- DBSCAN [34]: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) identifies high-density regions in the data space and groups points accordingly. Clusters are formed where the number of points in an ϵ-neighborhood exceeds a predefined threshold (*MinPts*):

  $$|\{x_j\in \text{Neighborhood}(x_i,\epsilon)\}|\geq \text{MinPts}$$

  DBSCAN is effective for detecting arbitrarily shaped clusters and handling noise, as it does not require predefining the number of clusters.

- Agglomerative Clustering [35]: A hierarchical clustering method that initially treats each data point as an individual cluster and iteratively merges clusters based on similarity. Various linkage criteria (single, complete, or average linkage) determine how clusters are merged, making it adaptable to different data structures.

- Spectral Clustering [36]: A graph-based clustering method that constructs an affinity matrix capturing pairwise similarities between data points. Eigenvalue decomposition is then applied to identify clusters. Spectral Clustering is particularly effective for non-convex data structures where traditional methods like K-Means may struggle.

- MeanShift [37]: A density-based clustering algorithm that iteratively shifts data points towards the nearest high-density region (mode). It does not require specifying the number of clusters in advance, making it adaptable to varying data distributions but computationally intensive for large datasets.

- Affinity Propagation [38]: An exemplar-based clustering algorithm that identifies representative points (exemplars) through a message-passing mechanism. Unlike K-Means, Affinity Propagation does not require specifying k in advance, making it highly adaptive to complex data structures.

- Fuzzy C-Means (FCM) [39]: A soft clustering technique where data points have varying degrees of membership to multiple clusters. The objective function is given by:

  $$J=\sum_{i=1}^{n}\sum_{j=1}^{k}u_{ij}^{m}\|x_i-\mu_j\|^2$$

  where u_{ij} represents the membership degree of x_i in cluster j, and $m>1$ controls the fuzziness level. FCM is effective when dealing with overlapping clusters.

These clustering techniques provide valuable information on the structure of the dataset, allowing the adaptive ensemble model to tailor its configurations to the distinct properties of each cluster. In the context of cardiovascular disease detection, these methods help uncover subgroups of patients that may correspond to varying risk profiles or stages of the disease.

Diverse base classifiers are employed, including RandomForestClassifier, SVC, KNeighborsClassifier, LogisticRegression, and NaiveBayes to capture different patterns in the dataset [40]. Each classifier offers distinct advantages: Random Forest leverages multiple decision trees for robust predictions, SVC maximizes the margin between classes using support vectors, k-NN classifies based on similarity measures, and Logistic Regression estimates the probability of binary classification as follows:

$$P(y=1|x)=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\cdots+\beta_px_p)}}$$

To maximize predictive performance, hyperparameter tuning is conducted using `GridSearchCV`, ensuring each model operates at its optimal configuration for the detection of cardiovascular diseases.

E. Ensemble Models: Static and Adaptive Configurations

To enhance model robustness and accuracy, ensemble methods are employed to combine predictions from multiple classifiers. Both static and adaptive configurations are considered, as outlined in Table II:

1) Static ensemble: A voting classifier aggregates predictions from tuned base models using soft voting [41]:

$$y=\arg\max_{c}\sum_{i}P(y_i=c)$$

where $P(y_i=c)$ represents the probability assigned by classifier i to class c. This approach leverages the collective predictive power of multiple classifiers to improve accuracy.

2) Adaptive ensemble: This approach applies three configurations that dynamically select specific models based on clustering labels, adapting to distinct cluster-specific patterns. A stacking classifier is further introduced, where the outputs of base classifiers serve as input to a meta-classifier, refining the final prediction.

These ensemble strategies improve the accuracy of the overall prediction by integrating the insights of multiple models, making them particularly effective for the detection of cardiovascular disease in diverse patient profiles.

F. Cross-Validation and Performance Evaluation

To ensure a rigorous evaluation, a 10-fold cross-validation is performed, preserving class distribution across folds. This stratified validation provides a reliable assessment of model generalization [42]. Key performance metrics: accuracy, precision, recall, F1 score, and AUC are calculated to evaluate the detection efficacy of each model:

- Accuracy: Measures the overall accuracy of the model.
- Precision: Reflects the reliability of positive predictions.
- Recall: Measures the sensitivity to actual positive cases.
- F1-score: Balances precision and recall.
- AUC: Evaluates the discriminative ability of the model.

This stage ensures a comprehensive evaluation of each model's ability to detect cardiovascular disease accurately and reliably.

G. Adaptive Ensemble Selection

Adaptive ensemble selection leverages clustering labels to dynamically tailor ensemble configurations for each identified cluster. By matching clusters with the most suitable ensemble models, this approach effectively captures variations within the dataset. This adaptability improves predictive accuracy by optimizing model selection for different subgroups of patients. In addition, it improves interpretability by providing information on the variability of the predictions in groups, supporting a more personalized and reliable approach to the detection of cardiovascular disease.

H. Result Aggregation and Analysis

Upon completing cross-validation, the performance metrics for each model configuration are averaged and analyzed. This aggregation identifies the configurations that achieve the best balance across key evaluation criteria, including accuracy, precision, recall, F1-score, and AUC. By highlighting the most effective models for the detection of cardiovascular disease, these insights provide an evidence-based assessment of predictive performance.

To facilitate detailed comparisons, the results are stored in a CSV file, allowing further analysis and evaluation. This structured approach supports a comprehensive assessment of the effectiveness of the adaptive ensemble system in improving disease detection accuracy.

I. Performance Measures

In assessing the effectiveness of classification models for the detection of cardiovascular disease, key performance metrics are evaluated: accuracy, precision, recall, F1 score and AUC-ROC [43]. These metrics provide a comprehensive view of each model's ability to classify instances correctly, balance detection between different health statuses, and maintain robust performance across varying classification thresholds.

- Accuracy quantifies the proportion of correctly classified instances in the detection of cardiovascular disease. It is computed as the ratio of correctly predicted cases (both positive and negative) to the total number of cases:

  $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

  where:
    - TP (True Positives): Correctly classified disease-positive cases.
    - TN (True Negatives): Correctly classified disease-negative cases.
    - FP (False Positives): Cases incorrectly classified as disease-positive.
    - FN (False Negatives): Cases incorrectly classified as disease negative.

  Accuracy provides an overall assessment of the correctness of the model in identifying both disease and non-disease cases.
- Precision measures the reliability of the model in identifying true disease cases among those classified as disease-positive, minimizing false positives. It is defined as:

  $$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

  A high-precision score indicates that, when the model predicts a positive case, it is likely correct.
- Recall (or sensitivity) evaluates the proportion of actual disease cases correctly identified by the model. This metric is particularly crucial in medical diagnostics, where missing actual disease cases (false negatives) can have serious consequences. Recall is computed as:

  $$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

  In the detection of cardiovascular disease, a high recall score ensures that most cases of disease are correctly identified.
- F1 Score provides a balanced measure of precision and recall. As the harmonic mean of these two metrics, it is particularly useful when both aspects are equally important. The F1 score is calculated as:

  $$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

  A high F1 score indicates that the model achieves a good balance between correctly identifying disease cases and minimizing false positives.

TABLE II. CLUSTER CONFIGURATION OF ENSEMBLE MODELS

Ensemble Type	Configuration	Description
Static Ensemble	Voting (Soft)	Combines predictions from `RandomForest` (RF), `SVC`, `k-NN`, and `Logistic Regression` (LR) using soft voting. This ensemble aggregates the predictions of each classifier and averages the probabilities to improve robustness and accuracy across the entire dataset.
Adaptive Ensemble 1	Voting (Soft)	Combines `RandomForest` (RF) and `Logistic Regression` (LR) using soft voting. This configuration adapts to clusters where tree-based and linear models best capture the underlying patterns.
Adaptive Ensemble 2	Voting (Soft)	Combines `SVC` and `k-NN` using soft voting. This configuration is applied to clusters that may benefit from both margin-based and instance-based classification techniques.
Adaptive Ensemble 3	Stacking (with Logistic Regression meta-classifier)	Integrates `RandomForest` (RF), `SVC`, and `k-NN`, with `Logistic Regression` as meta-classifier. The metaclassifier learns from the base classifiers' predictions, adapting to clusters where combined outputs from tree, margin, and instance-based models are beneficial.

- AUC-ROC (Area Under the Receiver Operating Characteristic Curve) evaluates the ability of the model to differentiate between disease and non-disease cases in varying classification thresholds. It is computed as the area under the ROC curve, which plots the true positive rate (recall) against the false positive rate:

  $$\text{AUC-ROC} = \int_{-\infty}^{+\infty} \text{TPR}(x)\, d\text{FPR}(x) \tag{5}$$

  where:
  - TPR (True Positive Rate) corresponds to the recall.
  - FPR (False Positive Rate) represents the proportion of non-disease cases incorrectly classified as disease positive.

  A higher AUC-ROC score indicates superior overall performance in distinguishing between disease and non-disease cases.

These performance metrics collectively provide a rigorous evaluation framework, helping identify the most effective model configurations for the detection of cardiovascular disease based on the data set and research objectives.

IV. RESULTS AND ANALYSIS

This section presents a comprehensive analysis of the results obtained from the evaluation of individual classifiers and ensemble methods under various clustering and feature selection configurations. The primary objective is to evaluate the performance of classifiers both independently and within adaptive and static ensemble frameworks.

The evaluation is carried out using key performance metrics, including accuracy, precision, recall, F1 score, and AUC, to determine the effectiveness of each approach in the detection of cardiovascular disease. The analysis provides insights into how different ensemble strategies and clustering techniques influence model performance, highlighting the most effective configurations.

A. Individual Classifiers Results

Table III presents a summary of the performance of individual classifiers under different feature selection methods. The evaluation reveals several key trends in classifier performance in various feature selection strategies.

In general, using all features (denoted as "All features") resulted in consistently strong performance across classifiers. In particular, Naive Bayes (NB) achieved the highest overall metrics, with an accuracy of 83%, precision of 85%, recall of 83%, F1 score of 83%, and AUC of 83%. This suggests that NB performs robustly when provided with the full feature set, likely due to its probabilistic nature and ability to handle redundant features effectively.

Support Vector Classifier (SVC) and Logistic Regression (LR) also exhibited stable performance across feature selection methods, with only slight variations. However, SVC demonstrated notable improvements with Chi-Squared feature selection, achieving the highest accuracy (84%) while maintaining strong scores in other metrics. This suggests that chi-square selection improves the ability of SVC to capture relevant patterns while reducing noise.

Additionally, Chi-Squared feature selection benefited K-Nearest Neighbors (KNN), improving both precision and recall compared to the full feature set. This improvement may indicate that the Chi-square selection aligns well with the neighborhood-based approach of KNN, likely by eliminating irrelevant or less discriminative features, thus refining similarity-based classification.

Overall, these results highlight the impact of feature selection on model performance, with Chi-Squared emerging as a particularly beneficial method to improve certain classifiers while maintaining overall predictive effectiveness.

In contrast, the selection of mutual information features (MutualInfo) produced mixed results between the classifiers. Naïve Bayes (NB) continued to perform well, maintaining high precision, recall, and AUC, demonstrating its resilience to feature reduction. However, MutualInfo negatively impacted Random Forest (RF), as indicated by a drop in accuracy (77%), precision (78%), recall (77%), F1 score (77%), and AUC (76%). This suggests that RF may rely on a wider set of features for optimal performance, as feature reduction could limit its ability to leverage multiple informative attributes.

Similarly, K-Nearest Neighbors (KNN) exhibited a decrease in the F1 score and AUC under MutualInfo, implying that, like RF, it benefits less from this feature selection method. The performance reduction may be due to the nature of KNN, which depends on distance-based comparisons, making it more sensitive to the availability of relevant features.

Overall, the results suggest that Naïve Bayes and Support Vector Classifier (SVC) exhibit more stable and resilient performance across feature selection methods, with SVC particularly excelling under Chi-Squared selection. In contrast, RF and KNN displayed greater sensitivity to feature selection,

TABLE III. 10-FOLD CROSS-VALIDATION RESULTS FOR DIFFERENT CLASSIFIERS UNDER VARIOUS FEATURE SELECTION METHODS

Classifier	Feature Selection	Accuracy	Precision	Recall	F1 Score	AUC
RF	All features	81.00%	82.00%	81.00%	81.00%	81.00%
SVC	All features	82.00%	83.00%	82.00%	82.00%	82.00%
KNN	All features	81.00%	82.00%	81.00%	80.00%	80.00%
LR	All features	82.00%	83.00%	82.00%	82.00%	82.00%
NB	All features	83.00%	85.00%	83.00%	83.00%	83.00%
RF	ChiSquared	80.00%	80.00%	80.00%	80.00%	80.00%
SVC	ChiSquared	84.00%	84.00%	84.00%	84.00%	83.00%
KNN	ChiSquared	82.00%	83.00%	82.00%	82.00%	82.00%
LR	ChiSquared	82.00%	83.00%	82.00%	82.00%	81.00%
NB	ChiSquared	81.00%	82.00%	81.00%	81.00%	81.00%
RF	MutualInfo	77.00%	78.00%	77.00%	77.00%	76.00%
SVC	MutualInfo	82.00%	83.00%	82.00%	82.00%	81.00%
KNN	MutualInfo	80.00%	81.00%	80.00%	79.00%	79.00%
LR	MutualInfo	81.00%	82.00%	81.00%	81.00%	80.00%
NB	MutualInfo	83.00%	84.00%	83.00%	83.00%	83.00%

especially under MutualInfo, indicating their preference for a larger set of features to maintain balanced precision, recall, and discriminatory power.

These findings emphasize the importance of selecting appropriate feature selection methods to optimize classifier performance. Specifically, Chi-Squared selection appears particularly beneficial for SVC and Logistic Regression (LR), whereas utilizing all features might yield the best results for NB.

B. Ensemble Methods Results

1) Performance analysis of adaptive and static ensemble selection: This section provides a comprehensive evaluation of the performance of adaptive and static ensemble selection methods, emphasizing the advantages of adaptive configurations, particularly when combined with various clustering techniques, across all evaluated metrics. The reported results are based on 10-fold cross-validation averages for each clustering method, with values averaged across two ensemble selectors, Multi-Layer Perceptron (MLP) and Decision Tree. Notably, the "All Features" feature selection case refers to configurations where all available features are used without reduction. By evaluating classifier performance across multiple folds and ensemble configurations, this analysis ensures a robust assessment of model stability and effectiveness.

Across all feature selection methods, adaptive ensemble selection consistently outperforms static ensemble, demonstrating its effectiveness in leveraging the underlying data structure. The static ensemble, which aggregates classifiers without considering data clusters, serves as a baseline and exhibits relatively lower performance across all metrics. Specifically, the static ensemble records lower average accuracy (80. 60% to 82. 81%), precision (81. 09% to 83. 33%), recall (80. 60% to 82. 81%), F1 score (80. 38% to 82. 57%) and AUC (79. 89% to 81. 99%). These results underscore the benefits of adaptive ensemble methods that dynamically adjust classifier configurations based on specific data clusters.

a) Impact of clustering on accuracy: Focusing on accuracy, the adaptive ensemble selection method demonstrates significantly higher performance, as illustrated in Fig. 1, particularly when combined with density-based clustering techniques such as MeanShift and DBSCAN. Under the "All features" setting, these clustering techniques achieve up to 95.57% and 93.35% accuracy, respectively. These results suggest that density-based clustering effectively captures natural groupings in the data, leading to more precise classifications within each cluster.

Even with Chi-Squared feature reduction, adaptive ensemble selection combined with Agglomerative and MeanShift clustering achieves an accuracy of 91.18%, indicating that these clustering methods retain essential information despite the reduced feature set. Furthermore, under Mutual Information feature selection, MeanShift clustering attains the highest accuracy (87.21%), demonstrating its robustness even when the feature set is limited to five selected attributes.

b) Impact on precision and recall: Turning to precision (Fig. 2), which measures the model's ability to minimize false positives, adaptive ensembles using MeanShift clustering achieve the highest precision (95. 70%) under the selection of "All features". DBSCAN and Affinity Propagation also demonstrate strong precision results, achieving up to 93 52%, highlighting the effectiveness of density-based and affinity-based clustering in improving classification reliability. Under Chi-Squared feature selection, Agglomerative and MeanShift clustering achieve the highest precision (91.48%), suggesting their ability to retain relevant features for accurate positive classifications.

Similarly, in terms of recall (Fig. 3), adaptive ensemble selection demonstrates an advantage, particularly when combined with MeanShift (95.57%) and DBSCAN (93.35%) clustering under the selection "All features". These results indicate that these clustering techniques allow the model to capture relevant patterns within the data, leading to more true positives. Even with feature reduction through chi-squared selection, Agglomerative and MeanShift clustering maintain a high recall (91. 18%), further confirming their ability to retain key features necessary to correctly identify positive cases.

c) Impact on F1 Score and AUC: Examining the F1 score (Fig. 4), which balances precision and recall, adaptive ensemble selection again outperforms static ensemble. The highest F1 score is observed in adaptive ensembles using MeanShift (95.54%) and DBSCAN (93.32%) clustering methods with the "All features" selection. With Chi-Squared selection, Agglomerative and MeanShift clustering maintain high F1 scores (91.12%), demonstrating their ability to preserve critical features for balanced classification performance despite feature reduction.

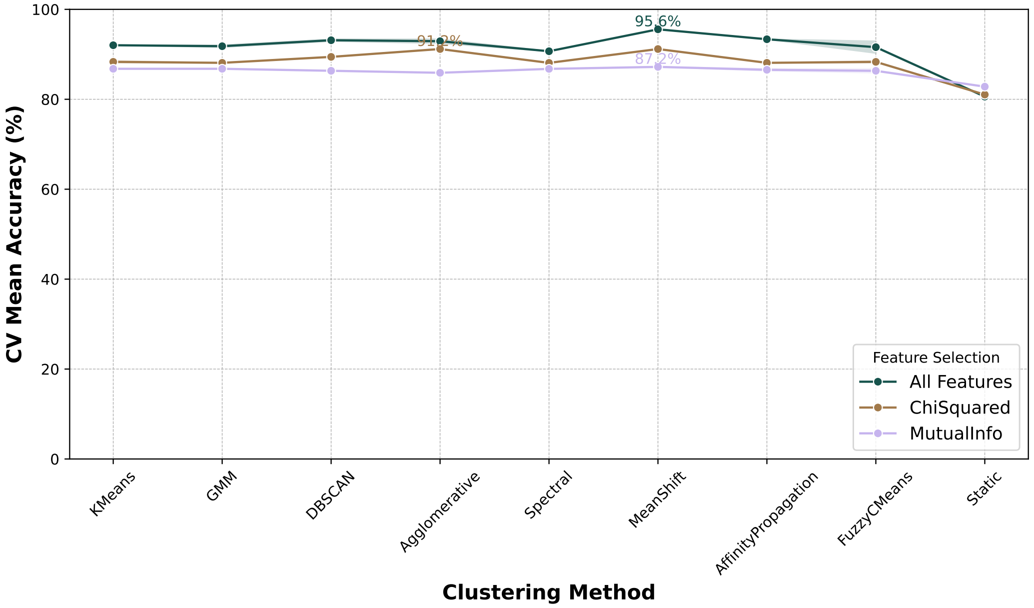

Fig. 1. CV Mean recall across different clustering and feature selection methods.

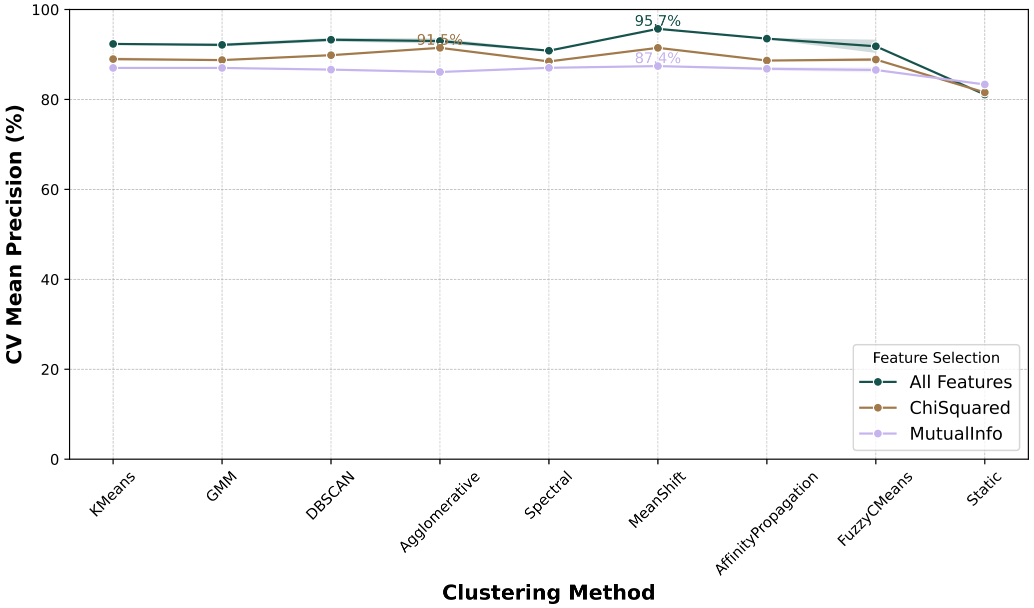

Fig. 2. CV Mean Precision across different clustering and feature selection methods.

Finally, considering AUC (Fig. 5), which assesses the model's ability to distinguish between classes, adaptive ensembles again achieve the highest scores. MeanShift clustering reaches the highest AUC (95.22%) under the “All features” selection, indicating strong discriminatory power and class separation. DBSCAN and Affinity Propagation clustering also demonstrate high AUC values (92.95%), reinforcing their role in improving class separation. Even under Chi-Squared feature selection, Agglomerative and MeanShift clustering achieve the highest AUC (90.62%), further validating their effectiveness in class discrimination.

d) Summary and key insights: In summary, adaptive ensemble selection shows consistent improvements in all metrics compared to the static ensemble approach. The ability of adaptive methods to dynamically adjust model selection based on cluster characteristics yields substantial performance gains, particularly when paired with MeanShift, DBSCAN, and Affinity Propagation clustering. These clustering techniques allow the adaptive ensemble to tailor model selection to different clusters, resulting in increased accuracy, precision, recall, F1 score, and AUC across different feature selection strategies.

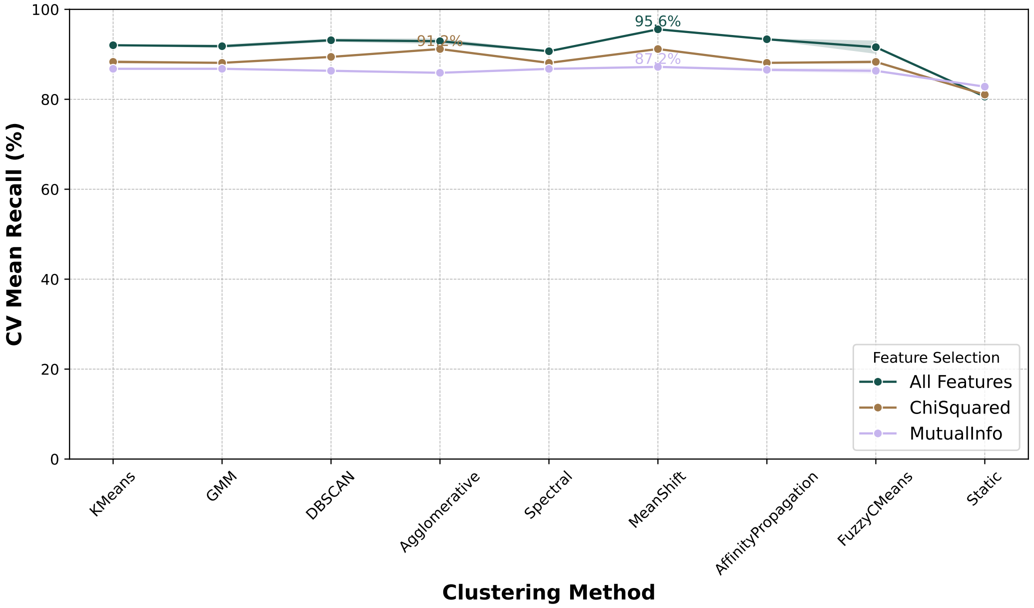

Fig. 3. CV Mean recall across different clustering and feature selection methods.

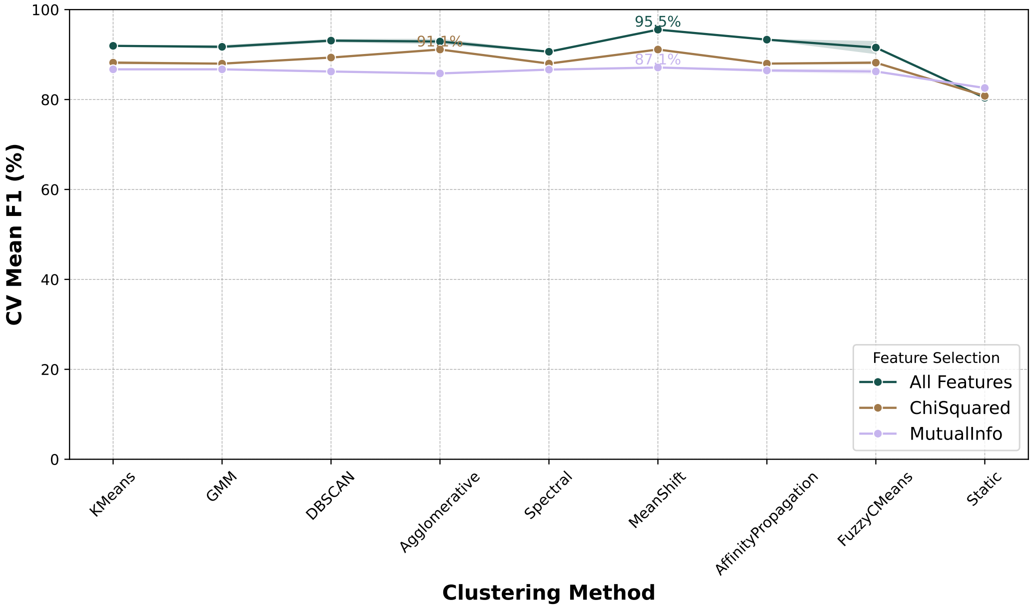

Fig. 4. CV Mean F1 Score across different clustering and feature selection methods.

In contrast, the static ensemble lacks this flexibility, showing consistently lower scores across all metrics. The greatest performance gaps are observed in accuracy and recall, where the adaptive ensemble's ability to capture distinct data patterns within groups allows significantly higher scores.

These findings highlight the effectiveness of adaptive ensemble selection in capturing nuanced data structures within clusters, offering superior performance over static methods. The results suggest that adaptive ensemble selection is particularly advantageous in complex datasets where clusters represent meaningful subgroups, as it allows the model to dynamically adjust to the intrinsic structure of the data. This capability provides a robust and precise classification framework with potential applications in medical diagnosis, risk assessment, and other high-stakes decision-making contexts.

2) Impact of ensemble selectors on adaptive ensemble performance: This subsection examines the influence of different ensemble selectors, namely Decision Tree and Multi-Layer Perceptron (MLP), on the performance of adaptive ensemble selection methods across various clustering techniques and feature selection strategies. The reported results are based on 10-fold cross-validation averages, as summarized in Table IV.

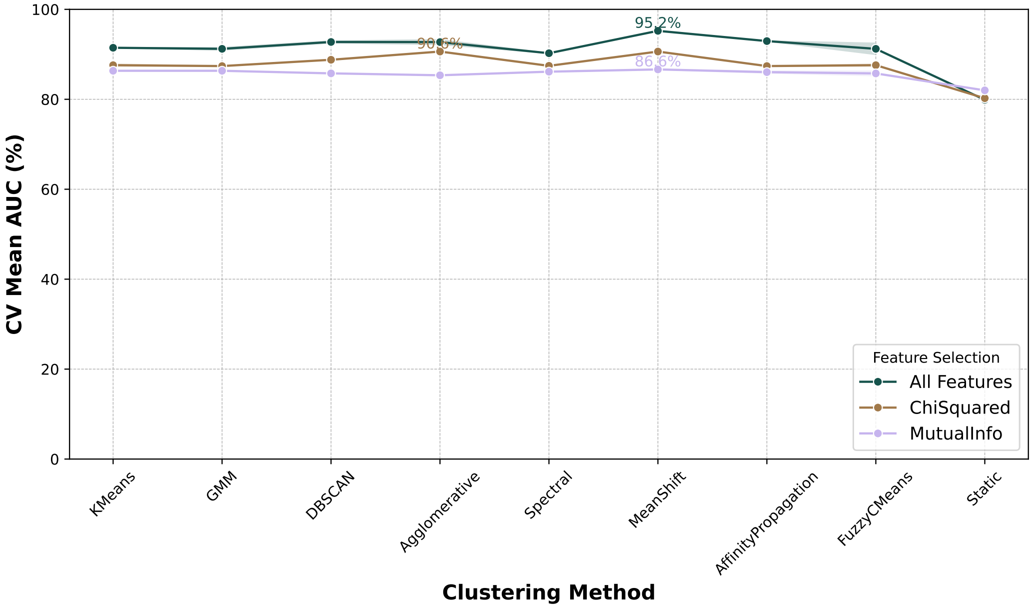

Fig. 5. CV Mean AUC across different clustering and feature selection methods.

a) Effect of ensemble selectors under "All Features" selection: Under the "All features" selection method, where all 13 features are used without reduction, the choice of ensemble selector has a generally minor impact on performance across clustering methods. However, some differences are notable.

- With DBSCAN clustering, the MLP ensemble selector achieves slightly higher accuracy (93.35%) compared to the Decision Tree selector (92.92%).
- For FuzzyCMeans clustering, MLP attains an accuracy of 92.91%, noticeably outperforming Decision Tree (90.28%).
- With MeanShift clustering, both ensemble selectors yield identical accuracy (95.57%), indicating that for this clustering technique, the choice of ensemble selector has negligible effect on performance.

This trend is consistently observed across all performance metrics, where MeanShift clustering achieves peak performance regardless of the ensemble selector. These observations suggest that while MLP may offer marginal gains with specific clustering techniques, the overall impact of the ensemble selector is minimal under the "All features" configuration.

b) Effect of ensemble selectors under ChiSquared feature selection: Under ChiSquared feature selection, which reduces the set of features to the five most relevant features based on correlation with the target variable, the impact of the ensemble selector remains varied:

- For K-Means clustering, MLP achieves a slightly higher accuracy (88.55%) than Decision Tree (88.11%), a trend that is mirrored in precision, recall, and F1 score metrics.
- Both Agglomerative and MeanShift clustering methods exhibit identical accuracy and F1 scores (91.18%) for both selectors, suggesting that these clustering techniques maintain consistent performance regardless of the ensemble selector.
- With FuzzyCMeans clustering, MLP marginally outperforms Decision Tree, achieving 88.54% accuracy compared to 88.11%.

These results indicate that while ChiSquared feature selection introduces some performance variations depending on the ensemble selector, the differences remain relatively small, with MLP offering minor improvements in certain clustering methods.

c) Effect of ensemble selectors under mutual information feature selection: Under Mutual Information feature selection, which retains the five most informative features based on their dependency on the target, the influence of the ensemble selector is even less pronounced:

- Across most clustering techniques, such as K-Means, GMM, and DBSCAN, both MLP and Decision Tree selectors achieve identical performance, with accuracy values of 86.78% for K-Means and GMM, and 86.34% for DBSCAN.
- MeanShift clustering, again, attains the highest accuracy (87.21%) for both MLP and Decision Tree selectors along with similar high precision and F1 scores.
- The only noticeable difference appears with FuzzyCMeans clustering, where MLP achieves slightly higher accuracy (86.77%) than Decision Tree (85.89%), as depicted in Table IV.

Overall, in Mutual Information selection, the ensemble selector has minimal impact on performance, and both selectors produce comparable results across clustering methods.

TABLE IV. 10-FOLD CROSS-VALIDATION RESULTS FOR DIFFERENT CLUSTERING AND ENSEMBLE SELECTOR CONFIGURATIONS UNDER VARIOUS FEATURE SELECTION METHODS

Feature Selection	Clustering Method	Ensemble Selector	CV Mean Accuracy	CV Mean Precision	CV Mean Recall	CV Mean F1	CV Mean AUC
All features	KMeans	DecisionTree	92.02%	92.37%	92.02%	91.92%	91.45%
All features	KMeans	MLP	92.03%	92.31%	92.03%	91.95%	91.47%
All features	GMM	DecisionTree	91.58%	91.95%	91.58%	91.49%	90.97%
All features	GMM	MLP	92.03%	92.31%	92.03%	91.95%	91.47%
All features	DBSCAN	DecisionTree	92.92%	93.03%	92.92%	92.88%	92.55%
All features	DBSCAN	MLP	93.35%	93.52%	93.35%	93.32%	92.95%
All features	Agglomerative	DecisionTree	93.36%	93.40%	93.36%	93.35%	93.22%
All features	Agglomerative	MLP	92.47%	92.58%	92.47%	92.44%	92.22%
All features	Spectral	DecisionTree	90.69%	90.83%	90.69%	90.63%	90.25%
All features	Spectral	MLP	90.69%	90.83%	90.69%	90.63%	90.25%
All features	MeanShift	DecisionTree	95.57%	95.70%	95.57%	95.54%	95.22%
All features	MeanShift	MLP	95.57%	95.70%	95.57%	95.54%	95.22%
All features	AffinityPropagation	DecisionTree	93.35%	93.52%	93.35%	93.32%	92.95%
All features	AffinityPropagation	MLP	93.35%	93.52%	93.35%	93.32%	92.95%
All features	FuzzyCMeans	DecisionTree	90.28%	90.51%	90.28%	90.25%	90.00%
All features	FuzzyCMeans	MLP	92.91%	93.13%	92.91%	92.85%	92.45%
All features	Static	Static Voting Ensemble	80.60%	81.09%	80.60%	80.38%	79.89%
ChiSquared	KMeans	DecisionTree	88.11%	88.75%	88.11%	87.98%	87.39%
ChiSquared	KMeans	MLP	88.55%	89.20%	88.55%	88.42%	87.79%
ChiSquared	GMM	DecisionTree	88.11%	88.75%	88.11%	87.98%	87.39%
ChiSquared	GMM	MLP	88.11%	88.75%	88.11%	87.98%	87.39%
ChiSquared	DBSCAN	DecisionTree	89.43%	89.84%	89.43%	89.33%	88.79%
ChiSquared	DBSCAN	MLP	89.43%	89.84%	89.43%	89.33%	88.79%
ChiSquared	Agglomerative	DecisionTree	91.18%	91.48%	91.18%	91.12%	90.62%
ChiSquared	Agglomerative	MLP	91.18%	91.48%	91.18%	91.12%	90.62%
ChiSquared	Spectral	DecisionTree	88.10%	88.45%	88.10%	87.99%	87.44%
ChiSquared	Spectral	MLP	88.10%	88.45%	88.10%	87.99%	87.44%
ChiSquared	MeanShift	DecisionTree	91.18%	91.48%	91.18%	91.12%	90.62%
ChiSquared	MeanShift	MLP	91.18%	91.48%	91.18%	91.12%	90.62%
ChiSquared	AffinityPropagation	DecisionTree	88.11%	88.65%	88.11%	87.99%	87.39%
ChiSquared	AffinityPropagation	MLP	88.11%	88.65%	88.11%	87.99%	87.39%
ChiSquared	FuzzyCMeans	DecisionTree	88.11%	88.65%	88.11%	87.99%	87.39%
ChiSquared	FuzzyCMeans	MLP	88.54%	89.10%	88.54%	88.41%	87.79%
ChiSquared	Static	Static Voting Ensemble	81.04%	81.56%	81.04%	80.81%	80.24%
MutualInfo	KMeans	DecisionTree	86.78%	87.01%	86.78%	86.72%	86.34%
MutualInfo	KMeans	MLP	86.78%	87.01%	86.78%	86.72%	86.34%
MutualInfo	GMM	DecisionTree	86.78%	87.01%	86.78%	86.72%	86.34%
MutualInfo	GMM	MLP	86.78%	87.01%	86.78%	86.72%	86.34%
MutualInfo	DBSCAN	DecisionTree	86.34%	86.63%	86.34%	86.23%	85.77%
MutualInfo	DBSCAN	MLP	86.34%	86.63%	86.34%	86.23%	85.77%
MutualInfo	Agglomerative	DecisionTree	85.89%	86.10%	85.89%	85.80%	85.34%
MutualInfo	Agglomerative	MLP	85.89%	86.10%	85.89%	85.80%	85.34%
MutualInfo	Spectral	DecisionTree	86.76%	87.03%	86.76%	86.65%	86.14%
MutualInfo	Spectral	MLP	86.76%	87.03%	86.76%	86.65%	86.14%
MutualInfo	MeanShift	DecisionTree	87.21%	87.41%	87.21%	87.12%	86.64%
MutualInfo	MeanShift	MLP	87.21%	87.41%	87.21%	87.12%	86.64%
MutualInfo	AffinityPropagation	DecisionTree	86.77%	87.04%	86.77%	86.67%	86.24%
MutualInfo	AffinityPropagation	MLP	86.34%	86.60%	86.34%	86.25%	85.84%
MutualInfo	FuzzyCMeans	DecisionTree	85.89%	86.14%	85.89%	85.79%	85.27%
MutualInfo	FuzzyCMeans	MLP	86.77%	86.97%	86.77%	86.69%	86.24%
MutualInfo	Static	Static Voting Ensemble	82.81%	83.33%	82.81%	82.57%	81.99%

d) Comparison with static ensemble selection: When comparing adaptive ensemble selection with the static ensemble (which combines classifiers without clustering or dynamic selection), it is evident that adaptive configurations, regardless of the ensemble selector, consistently outperform the static case. The static ensemble's accuracy ranges from 80.60% to 82.81% across feature selection methods, significantly lower than the highest accuracy achieved by adaptive ensembles. This trend is consistent in all performance metrics, precision, recall, F1 score, and AUC, reinforcing the superiority of adaptive ensemble selection over static methods.

e) Summary and key insights: The results demonstrate that the choice of the ensemble selector (MLP or Decision Tree) has a relatively minor influence on the performance of adaptive ensemble selection. Although MLP offers slight advantages in specific configurations, particularly when combined with clustering methods such as DBSCAN and FuzzyCMeans, both ensemble selectors exhibit comparable high performance under the MeanShift clustering method.

These findings suggest that the main driver of improved performance in adaptive ensemble selection is the clustering method, while the ensemble selector plays a secondary role.

C. Comparison with Existing Methods

As shown in Table V, the proposed dynamic ensemble method significantly outperforms existing approaches on the Cleveland dataset. Achieving an accuracy of 95.57%, the proposed method surpasses the best reported static ensemble approach, which achieves an accuracy of 92.11% [27], by a margin of 3.46%.

In addition, other notable studies, such as [19] and [8], report lower accuracies of 88.70% and 87.78%, respectively, further strengthening the superior performance of the proposed

TABLE V. PERFORMANCE COMPARISON OF THE PROPOSED DYNAMIC ENSEMBLE METHOD WITH EXISTING APPROACHES ON THE CLEVELAND DATASET

Reference	**Year**	**Dataset**	**Ensemble Type**	**Accuracy (%)**
[8]	2019	Cleveland	Static	87.78
[16]	2019	Cleveland	Static	84.79
[18]	2020	Cleveland	Static	85.71
[21]	2020	Cleveland	Static	87.30
[23]	2020	Cleveland	Static	75–86
[22]	2021	Cleveland	Static	83.00
[24]	2021	Cleveland	Static	87.05
[15]	2022	Cleveland	Static	87.00
[19]	2022	Cleveland	Static	88.70
[27]	2024	Cleveland	Static	92.11
Proposed	**2025**	**Cleveland**	**Dynamic (MeanShift, MLP)**	**95.57**

approach. These findings validate the efficacy of the adaptive framework, particularly the integration of the MeanShift clustering method and MLP as the ensemble selector.

Importantly, the results highlight that the choice of clustering technique plays a pivotal role in enhancing the performance of the ensemble by forming more effective groups of base classifiers, which the dynamic selection mechanism then optimally leverages.

Thus, the proposed method represents a significant advancement in adaptive ensemble selection for the prediction of cardiovascular disease, achieving substantial improvements over state-of-the-art approaches in terms of predictive accuracy and classification performance.

V. CONCLUSION

Cardiovascular disease (CVD) poses a significant global health challenge, making early and accurate prediction essential for improving patient outcomes and reducing healthcare care burdens. In this study, an adaptive ensemble selection approach was proposed to improve CVD prediction by dynamically tailoring model configurations to distinct patient subgroups. By integrating various clustering techniques, such as K-Means, DBSCAN, and MeanShift, with feature selection methods, including Chi-Squared and Mutual Information, this approach aimed to improve predictive performance by adapting to the unique characteristics of each group of patients. Ensemble selectors were tested with both Multi-Layer Perceptron (MLP) and Decision Tree configurations to assess their effectiveness across different clustering strategies.

The findings indicated that adaptive ensemble selection consistently outperformed static ensemble in all key performance metrics, including accuracy, precision, recall, F1 score, and AUC. Specifically, the use of MeanShift and DBSCAN, combined with the retention of all characteristics, produced the highest accuracy, demonstrating the effectiveness of clustering based on density to capture meaningful patterns in patient data. These results highlight the advantages of adaptive ensemble selection in leveraging cluster-specific insights, particularly in complex, heterogeneous datasets where patient subgroups may differ in risk profiles or disease stages.

In summary, this study demonstrated that adaptive ensemble selection, particularly when paired with density-based clustering methods like MeanShift, holds substantial promise for personalized CVD prediction. By dynamically adjusting to the underlying data structure, this adaptive approach offers a scalable and robust solution for improving diagnostic accuracy in high-stakes medical applications. The results suggest that adaptive ensemble methods could serve as a valuable tool in personalized healthcare, allowing more targeted and effective interventions tailored to individual patient needs.

ACKNOWLEDGMENT

Machine learning training and evaluation have been performed using the Phoenix High Performance Computing facility at the American University of the Middle East, Kuwait.

REFERENCES

[1] S. Rajalakshmi and K. V. Madhav, “A collaborative prediction of presence of arrhythmia in human heart with electrocardiogram data using machine learning algorithms with analytics,” *Journal of Computer Science*, vol. 15, no. 2, pp. 278–287, Feb 2019.

[2] S. Hiriyannaiah, S. G M, K. M H M, and K. G. Srinivasa, “A comparative study and analysis of lstm deep neural networks for heartbeats classification,” *Health and Technology*, vol. 11, no. 3, pp. 663–671, May 2021.

[3] World Health Organization, “World health statistics,” 2024, available online: https://www.who.int/data/gho/publications/world-health-statistics (accessed on 1 October 2024).

[4] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, “Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals,” *Computers in Biology and Medicine*, vol. 94, pp. 19–26, 2018.

[5] P. Bizopoulos and D. Koutsouris, “Deep learning in cardiology,” *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 168–193, 2019.

[6] S. Kaur, J. Singla, L. Nkenyereye, S. Jha, D. Prashar, G. P. Joshi, S. El-Sappagh, M. S. Islam, and S. M. R. Islam, “Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives,” *IEEE Access*, vol. 8, pp. 228 049–228 069, 2020.

[7] M. M. Taye, “Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions,” *Computers*, vol. 12, no. 5, 2023.

[8] H. Al-Khazraji, A. R. Nasser, A. M. Hasan, A. K. Al Mhdawi, H. Al-Raweshidy, and A. J. Humaidi, “Aircraft engines remaining useful life prediction based on a hybrid model of autoencoder and deep belief network,” *IEEE Access*, vol. 10, pp. 82 156–82 163, 2022.

[9] M. Khalifa, M. Albadawy, and U. Iqbal, “Advancing clinical decision support: The role of artificial intelligence across six domains,” *Computer Methods and Programs in Biomedicine Update*, vol. 5, p. 100142, Jan 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666990024000090

[10] P. Rahman, A. Rifat, M. Chy, M. M. Khan, M. Masud, and S. Aljahdali, “Machine learning and artificial neural network for predicting heart failure risk.” *Computer Systems Science & Engineering*, vol. 44, no. 1, 2023.

[11] D. Asif, M. Bibi, M. S. Arif, and A. Mukheimer, “Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization,” *Algorithms*, vol. 16, no. 6, 2023.

[12] A. AlMohimeed, H. Saleh, S. Mostafa, R. M. A. Saad, and A. S. Talaat, “Cervical cancer diagnosis using stacked ensemble model and optimized feature selection: An explainable artificial intelligence approach,” *Computers*, vol. 12, no. 10, 2023.

[13] L. Miao and W. Wang, “Cardiovascular disease prediction based on soft voting ensemble model,” *Journal of Physics: Conference Series*, vol. 2504, no. 1, p. 012021, may 2023.

[14] V. Shorewala, “Early detection of coronary heart disease using ensemble techniques,” *Informatics in Medicine Unlocked*, vol. 26, p. 100655, 2021.

[15] V. Jain and K. L. Kashyap, “Multilayer hybrid ensemble machine learning model for analysis of covid-19 vaccine sentiments,” *Journal of Intelligent & Fuzzy Systems*, vol. 43, pp. 6307–6319, 2022, 5.

[16] F. A. Vellameeran and T. Brindha, "A new variant of deep belief network assisted with optimal feature selection for heart disease diagnosis using iot wearable medical devices," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 25, no. 4, pp. 387–411, 2022, pMID: 34311642. [Online]. Available: https://doi.org/10.1080/10255842.2021.1955360

[17] Srınıvasa Rao, B., "A new ensenble learning based optimal prediction model for cardiovascular diseases," *E3S Web Conf.*, vol. 309, p. 01007, 2021. [Online]. Available: https://doi.org/10.1051/e3sconf/202130901007

[18] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular disease detection using ensemble learning," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 5267498, 2022.

[19] B. Baranidharan, A. Pal, and P. Muruganandam, "Cardiovascular disease prediction based on ensemble technique enhanced using extra tree classifier for feature selection," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 3236–42, 2019.

[20] S. Diwan, G. S. Thakur, S. K. Sahu, M. Sahu, and N. K. Swamy, "Predicting heart diseases through feature selection and ensemble classifiers," *Journal of Physics: Conference Series*, vol. 2273, no. 1, p. 012027, may 2022.

[21] 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.

[22] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *BioMed Research International*, vol. 2020, no. 1, p. 9816142, 2020.

[23] X. Wenxin, "Heart disease prediction model based on model ensemble," in *2020 3rd international conference on artificial intelligence and big data (ICAIBD)*. IEEE, 2020, pp. 195–199.

[24] S. Bashir, A. A. Almazroi, S. Ashfaq, A. A. Almazroi, and F. H. Khan, "A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction," *IEEE Access*, vol. 9, pp. 130 805–130 822, 2021.

[25] I. Javid, A. K. Z. Alsaedi, and R. Ghazali, "Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.

[26] N. Harika, S. R. Swamy, and Nilima, "Artificial intelligence-based ensemble model for rapid prediction of heart disease," *SN Computer Science*, vol. 2, no. 6, p. 431, Aug 2021.

[27] A. E. Korial, I. I. Gorial, and A. J. Humaidi, "An improved ensemble-based cardiovascular disease detection system with chi-square feature selection," *Computers*, vol. 13, no. 6, 2024.

[28] S. A. Ali, B. Raza, A. K. Malik, A. R. Shahid, M. Faheem, H. Alquhayz, and Y. J. Kumar, "An optimally configured and improved deep belief network (oci-dbn) approach for heart disease prediction based on ruzzo–tompa and stacked genetic algorithm," *IEEE Access*, vol. 8, pp. 65 947–65 958, 2020.

[29] J. Vijayashree and H. Parveen Sultana, "Heart disease classification using hybridized ruzzo-tompa memetic based deep trained neocognitron neural network," *Health and Technology*, vol. 10, no. 1, pp. 207–216, Jan 2020.

[30] V. Rupapara, F. Rustam, A. Ishaq, E. Lee, and I. Ashraf, "Chi-square and pca based feature selection for diabetes detection with ensemble classifier," *Intell. Autom. Soft Comput*, vol. 36, no. 2, pp. 1931–1949, 2023.

[31] H. Zhou, X. Wang, and R. Zhu, "Feature selection based on mutual information with correlation coefficient," *Applied Intelligence*, vol. 52, no. 5, pp. 5457–5474, Mar 2022.

[32] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, 2023.

[33] E. Patel and D. S. Kushwaha, "Clustering cloud workloads: K-means vs gaussian mixture model," *Procedia Computer Science*, vol. 171, pp. 158–167, 2020, third International Conference on Computing and Network Communications (CoCoNet'19).

[34] O. Kulkarni and A. Burhanpurwala, "A survey of advancements in dbscan clustering algorithms for big data," in *2024 3rd International conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC)*, 2024, pp. 106–111.

[35] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," *WIREs Computational Statistics*, vol. 15, no. 3, p. e1597, 2023.

[36] J. Xie, W. Kong, S. Xia, G. Wang, and X. Gao, "An efficient spectral clustering algorithm based on granular-ball," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 9, pp. 9743–9753, 2023.

[37] R. SETYAWAN and G. C. PAMUJI, "Comparative study of k-means and mean shift clustering algorithms for waste data in west java province," *Journal of Engineering Science and Technology*, vol. 19, no. 3, pp. 869–879, 2024.

[38] Y. Wang, C. Tao, Z. Zhou, K. Lin, C. K. Law, and B. Yang, "Clustering algorithm for experimental datasets using global sensitivity-based affinity propagation (gsap)," *Combustion and Flame*, vol. 259, p. 113121, 2024.

[39] D. Krasnov, D. Davis, K. Malott, Y. Chen, X. Shi, and A. Wong, "Fuzzy c-means clustering: A review of applications in breast cancer detection," *Entropy*, vol. 25, no. 7, 2023.

[40] M. A. B. Al-Tarawneh, O. Al-irr, K. S. Al-Maaitah, H. Kanj, and W. H. F. Aly, "Enhancing fake news detection with word embedding: A machine learning and deep learning approach," *Computers*, vol. 13, no. 9, 2024.

[41] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," *Healthcare*, vol. 11, no. 12, 2023.

[42] M. T R, V. K. V, D. K. V, O. Geman, M. Margala, and M. Guduri, "The stratified k-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthcare Analytics*, vol. 4, p. 100247, 2023.

[43] G. ALMahadin, M. O. Hiari, A. H. Hussein, N. M. M. Turab, A. Alkhresheh, and M. A. B. Al-Tarawneh, "Performance evaluation of an intelligent and optimized machine learning framework for attack detection," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 14, no. 3, p. 358–371, Dec. 2022.