**RESEARCH ARTICLE**

# An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers

## D. YASO OMKARI[ID] AND KAREEMULLA SHAIK[ID]
School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India

Corresponding author: Kareemulla Shaik (kareemulla.shaik@vitap.ac.in)

**ABSTRACT** Cardiovascular problems have emerged as a significant concern, adversely impacting individuals across all age groups. Several recent research studies have used Machine learning (ML) techniques to design decision-making systems for the tremendous data in the medical sector. Although these works obtained promising results, most of the studies focused on small datasets. Since the size of the dataset affects algorithm performance, this study used two datasets, such as Kaggle's heart disease dataset of over 70,000 records and UCI's heart disease dataset of 1025 records. In addition to the old features the Pulse Pressure (PP), the Body Mass Index (BMI), and the Mean Arterial Pressure (MAP), three new features are introduced to improve the results. This paper proposes the TLV (Two-Layer Voting) model, which is an ensemble method of hard and soft voting. As part of layer 1, features are shortlisted by soft and hard voting using three statistical methods, including the ANOVA f-test, Chi-squared test, and Mutual Information. In layer 2, soft voting and hard voting performance are compared, which incorporates Multi-Layer Perceptron, Decision Tree, Support Vector Classifier, and Random Forest algorithms. Classification algorithms are hyper-tuned using the GridSearchCV method in the second layer. Using UCI's heart disease dataset and Kaggle's CVD dataset, the proposed TLV methodology with soft voting provided the highest accuracy of 99.03% and 88.09%, respectively. The proposed model significantly outperforms existing CAD disease prediction studies.

**INDEX TERMS** ANOVA f-test, Chi-squared test, decision tree, heart disease, random forest, support vector classifier, multi-layer perceptron.

## I. INTRODUCTION

Technology's profound impact on the world in recent years has brought about various changes that influence people's health. A sedentary lifestyle and increasing reliance on the virtual realm have exposed individuals to various health risks, including heart disease [1]. Information about the behavioural patterns that contribute to cardiovascular disease is contained in health records. A sedentary lifestyle leads to obstacle formation in the artery walls. Unhealthy eating habits with high saturated fats and cholesterol can accelerate the progression of cardiovascular issues. Smoking damages the blood vessels and increases the risk of developing this condition. High

blood pressure and hypertension, as well as being overweight, are additional risk factors. Individuals with diabetes and insulin resistance are also more susceptible to heart diseases. High-stress levels and certain chronic conditions such as metabolic syndrome and kidney disease can further exacerbate the risk of developing heart disease [2], [3]. Numerous tests are conducted before the diagnosis of the disease such as auscultation, untreated raised or low blood pressure, bad cholesterol, electrocardiogram, and blood sugar analysis [4].

With technological advancements impacting every facet of medical care, Machine Learning (ML), is a vital tool for identifying cardiovascular diseases. ML can derive valuable insights from extensive datasets generated in the course of daily healthcare activities, thereby holding immense potential for significant contributions to the field of healthcare. Due to

---

The associate editor coordinating the review of this manuscript and approving it for publication was Abdallah Kassem[ID].

existing methods' limitations in accuracy and computational power, researchers are developing new ways to estimate heart diseases in their starting stages.

The majority of risk prediction algorithms concentrate on a limited number of risk factors. As a result of complex interactions between risk factors, these prediction systems struggle to perform well [5].

Numerous experiments have analyzed the ML and Deep Learning (DL) models for diagnosing cardiac diseases accurately [6]. An effective disease prediction may be hindered by choosing relevant features [7], a limited number of medical datasets, and a lack of in-depth analysis of risk factors. In most cases, these models are trained and evaluated using publicly available datasets. Patients' disease status and associated risk factors are included in these datasets.

Most of the studies used various datasets from the UCI repository and Kaggle websites. The Cleveland dataset [8], [9], the UCI heart disease dataset [10], and the Z-Alizadeh Sani dataset [11] are the most well-known datasets for Coronary Artery Disease (CAD) detection. Many researchers are working with small datasets, but the volume of medical data created now is enormous. The working of the ML strategies often degrades when the dataset is increased. Recent and large data should be analyzed and find the best one among ML models. To look into this issue, we are utilizing a very large heart disease dataset obtained from Kaggle, which contains over 70,000 records. In this paper, we aim to develop a model that enhances early-stage cardiovascular prediction.

We aim to accomplish the following objectives in this paper: implement a highly efficient cardiovascular disease classification system based on several ML approaches. In this paper, we have used two cardiovascular datasets namely Kaggle's heart disease dataset, and UCI's heart disease dataset for the evaluation of classification algorithms. To elevate the effectiveness of the proposed methodology by removing insignificant features we carried out the most significant feature selection using hard voting of three statistical methods namely ANOVA f-test, Chi-squared test, and Mutual Information. To achieve better results, the Random Forest, the Decision Tree, the Support Vector Classifier, and the Multi-Layer Perceptron are hyper-tuned in the second layer by using the GridsearchCV method. As a result, we could enhance the results of the proposed model by two-layered hard and soft voting with the hyperparameter tuning on both UCI's heart disease dataset and Kaggle's heart disease datasets.

### A. CONTRIBUTION
1. The research employs a rigorous feature selection process by utilizing hard voting of three statistical methods (ANOVA f-test, Chi-squared test, and Mutual Information) to identify the most significant features, thereby improving the effectiveness of the classification system.
2. To further enhance the performance of the proposed methodology, the research conducts hyperparameter tuning of key classification algorithms (Random Forest,

Decision Tree, Support Vector Classifier, and Multi-Layer Perceptron) using the GridsearchCV method, thereby optimizing the model's predictive capabilities.
3. By incorporating a two-layered approach of hard and soft voting, coupled with hyperparameter tuning, the research achieves improved results on both UCI's heart disease dataset and Kaggle's heart disease dataset, demonstrating the effectiveness and reliability of the proposed classification system.
4. The research work introduces a novel approach to classifying cardiovascular diseases using machine learning approaches, aiming to improve accuracy and efficiency in disease prediction.
5. By using both Kaggle's heart disease dataset and UCI's heart disease dataset, the research ensures a robust evaluation of classification algorithms, thereby enhancing the reliability and generalizability of the proposed methodology.

The organization of this research work is as follows: Section II gives information about related studies which are close to our work. Section III describes the dataset description, feature selection methods, and several ML and DL techniques used for the classification and also discusses the various measures used for the performance of the system. Section IV describe the preprocessing, and discussion on the results from this study, and the determinations are presented in Section V.

## II. RELATED WORKS
Numerous ML and DL techniques have been applied to disease prediction systems in the medical field. Gradient descent optimization [12], Deep neural networks [13], bagging ensemble methods [14], XGBoost [15], J48 [16], Random Forest [17], and Decision tree [18] were widely used in the classification of disease. A hybrid approach was created by Mohan et al. [19] that demonstrated an innovative way to extract necessary properties from data to understand and classify according to the vital patterns by using ML classifiers. In a study by Gárate-Escamil et al. [20] Hungarian-Cleveland datasets were used for classification. ML algorithms were used for the classification of heart disease, and PCA was used to reduce dimensionality and select features. According to Bharti et al. [21] the prediction accuracy was enhanced by uniting the Isolation Forest classifier with DL algorithms.

The classification of cardiac disease is based on various publicly available data sets. The authors [22], used an ensemble method to prediction of heart disease. The performance of the classifiers could be significantly enhanced using bagging and boosting methods. For hybrid model development, they used majority voting of Nave Bayes, Bayes Net, C 4.5, Multilayer Perceptron, PART, and Random Forest (RF) classifiers. With the designed model, an accuracy of 85.48% was obtained. Mienye et al. [23] proposed a prediction model for heart disease that uses the mean-based splitting approach. Afterward, a weighted classifier generated 93% and 91% classification accuracy on the Cleveland and Framingham

tests, respectively. Tama et al. [24] suggested a two-tier ensemble-based coronary disease (CHD) detection model. In this experiment, three various ensemble learners were used: gradient boosting, random forest, and extreme gradient boosting. According to the proposed model, accuracy, F1, and AUC values are 98.13%, 96.6%, and 98.7%, respectively.

Since most of the studies use small-sized datasets, this study uses a very large cardiovascular dataset from Kaggle. In [25], Maiga et al. analyzed the various risk factors for cardiovascular disease prediction. In this study, authors normalized data using a min-max scalar and applied k-fold cross-validation. Random Forest (RF), NB, Logistic Regression (LR), and K Nearest Neighbors (KNN) are applied to the dataset. This has been observed that RF performs well and gives the highest accuracy among all algorithms. However, this study needs to be enhanced with several feature selection methods to improve the results. Some of the studies more concentrated on feature selection methods for accurate results.

In a study by [26], the authors applied feature extraction techniques, namely embedded feature selection based on embedded methods, filter methods which are based on statistical tests, and wrapper methods. After extracting the feature subset using these methods, they evaluated the performance of ML algorithms. Notably, with the reduced feature set, XGBoost demonstrated the highest results compared to the others.

When it comes to forecasting heart disease, many studies have shown that ensemble approaches have proven to be extremely effective. Shorewala [27] designed a system with ensemble modeling techniques. Correlations between features were analyzed using the Pearson coefficient in this study. The LASSO method was applied for attribute extraction. Bagging, boosting, and stacking techniques were applied for detailed analysis. Among all, stacking is proven as the most effective model compared with base models. Further, tuning parameters can improve performance and various cross-validation methods for the model. Another study [28] proved the effectiveness of ensemble algorithms by comparing the performance of base models and ensemble techniques like stacking, bagging, and boosting. The features are selected by using the LASSO technique. Authors found that nearly 2% accuracy is improved by bagging models.

Many authors combine ML, statistics, and database systems to discover patterns in large datasets through Data Mining in [29]. In the work by B. Martins et al., the Cross Industry Standard Process was employed with algorithms, including decision tree (DT), RF, DL, optimized DT, and rule induction (RI). It was observed that the optimized DT outperformed all other algorithms during the experimental analysis.

The most accurate performance may be achieved by applying cross-validation to the algorithm. Some studies analyzed thoroughly the working of ML and DL models on small and large-sized datasets. In [30] authors used two datasets Cleveland (303 records) and Kaggle's heart disease (70000 records) dataset and this study analyzed the impact of various cross-validation techniques namely stratified k-fold cross-validation, hold-out, repeated random, and k-fold. On the Kaggle dataset neural networks with hold-out cross-validation and the Cleveland dataset RF with repeated random method achieved the highest accuracy.

In another study [31], authors used two datasets namely the Kaggle heart disease dataset (70000 records) and the UCI arrhythmia dataset (452 records). RF, bagging, Gradient boosting, and Extra tree algorithms are applied to the large dataset for model evaluation, and RF and Gradient boosting achieved the highest accuracy among the remaining algorithms. All ensemble techniques perform well on a small dataset. Furthermore, parameter optimization is a requisite in this paper. The risk factors in a large dataset should be analyzed in depth to improve the model's working.

In [32] authors applied various feature selection methods namely ANOVA f-test and f-classify methods. In this study, authors analyzed the performance of ML algorithms with the top 3, 8, and 12 features. SVM with the top 3 features and RF with the top 8 and 12 attributes achieved the highest accuracy among the remaining ML algorithms. In this study, the authors concentrated only on feature selection and the enhancement of results can be done by applying hyperparameter optimization and cross-validation techniques [33].

Drawing from the sources reviewed above, it becomes evident that feature selection is a crucial step in enhancing results. Furthermore, parameter optimization plays a pivotal role in elevating the model's overall working. To improve the results, the new attributes are created by feature engineering methods and incorporated into our dataset. Several feature extraction algorithms are applied to find the best features and hyperparameter tunning is done for enhancing the performance of the model.

### A. RESEARCH GAP ANALYSIS

The research gap in disease prediction models from Table 1, particularly in cardiovascular disease, underscores the need for advancements in feature selection, parameter optimization, and cross-validation techniques. While existing literature acknowledges the importance of feature selection methods like embedded feature selection and filter methods, there's room for exploring additional techniques. Similarly, more sophisticated parameter tuning methods are required to optimize machine learning algorithms effectively. Limited exploration of cross-validation methods, including stratified k-fold and repeated random, calls for a comprehensive evaluation to assess their impact on model generalization. Furthermore, enhancing model performance on large datasets remains a significant gap, necessitating thorough analyses and identification of key risk factors. Addressing these gaps can lead to more accurate and reliable disease prediction models in healthcare.

This research work addresses the research gap by acknowledging the limitations of previous studies, which

**TABLE 1.** Literature survey studies.

| Ref. No | Model Used | Advantages | Disadvantages | Research Gap |
|---|---|---|---|---|
| [12] | Gradient Descent Optimization | Efficient convergence to local minimums | Sensitive to learning rates | Feature selection impact on optimization process |
| [13] | Deep Neural Networks | Ability to capture complex patterns | Prone to overfitting | Integration with other algorithms for improvement |
| [14] | Bagging Ensemble Methods | Reduces variance and improves stability | Computationally intensive | Enhanced performance with large datasets |
| [15] | XGBoost | High predictive accuracy | Sensitive to parameter tuning | Integration with other algorithms for optimization |
| [16] | J48 (C4.5 Decision Tree) | Interpretability and simplicity | Tends to overfit with complex datasets | Ensemble with other classifiers for accuracy |
| [17] | Random Forest | Handles large datasets efficiently | Prone to overfitting | Investigation into feature importance |
| [18] | Decision Tree | Intuitive representation of decisions | Sensitive to data variations | Integration with ensemble methods for improvement |
| [19] | Hybrid Approach | Utilizes strengths of multiple algorithms | Complexity in combining different models | Evaluation of hybrid approach with large datasets |
| [20] | Principal Component Analysis (PCA) | Dimensionality reduction | Loss of interpretability | Integration with classification algorithms |
| [21] | Isolation Forest with DL Algorithms | Anomaly detection in datasets | Performance impact on large datasets | Investigation into combination impact on accuracy |
| [22] | Ensemble Method | Enhanced predictive accuracy | Computational complexity | Optimization of ensemble methods for efficiency |
| [23] | Mean-Based Splitting Approach | Simple approach for prediction | Limited scalability | Investigation into alternative splitting methods |
| [24] | Ensemble-Based Detection Model | High accuracy and performance | Model complexity and tuning requirements | Evaluation of ensemble methods with diverse datasets |
| [25] | Various ML Algorithms | Flexibility and adaptability | Performance impact on feature selection | Integration with feature engineering for improvement |
| [26] | Feature Extraction Techniques | Improved model performance | Computational overhead | feature extraction exploration |
| [27] | Ensemble Modeling Techniques | Improved accuracy through combination | Increased complexity and tuning requirements | Optimization of ensemble models for efficiency |
| [28] | Ensemble Algorithms | Enhanced predictive performance | Complexity in combining models | Investigation into optimal |

**TABLE 1.** *(Continued.)* Literature survey studies.

| | | | | ensemble configurations |
|---|---|---|---|---|
| [29] | Decision Tree, Random Forest, DL | Effective pattern discovery in large data | Complexity in algorithm selection | Evaluation of ensemble approach with diverse data |
| [30] | Various Cross-Validation Techniques | Enhanced model evaluation | Computational overhead | Optimization of cross-validation techniques |
| [31] | RF, Bagging, Gradient Boosting | High accuracy on large datasets | Model complexity and parameter tuning | Investigation into optimal ensemble configurations |
| [32] | Feature Selection Methods | Improved model efficiency and performance | Dimensionality reduction | Integration with hyperparameter optimization |
| [33] | Feature Engineering Methods | Enhanced model effectiveness | Complexity in feature creation | Investigation into novel feature engineering methods |

datasets, including Kaggle's heart disease dataset with over 70,000 records and UCI's heart disease dataset with 1025 records. Additionally, the introduction of three new features, Pulse Pressure (PP), Body Mass Index (BMI), and Mean Arterial Pressure (MAP), aims to enhance the predictive accuracy of the models. The proposed Two-Layer Voting (TLV) model, employing ensemble methods of hard and soft voting, demonstrates a novel approach to feature selection and algorithm comparison. By integrating statistical methods like ANOVA f-test, Chi-squared test, and Mutual Information in the first layer and fine-tuning classification algorithms in the second layer, the TLV model achieves superior accuracy rates compared to existing prediction studies, thus bridging the research gap and offering advancements in cardiovascular disease prediction.

## III. MATERIALS AND METHODS

ML models rely heavily on feature selection or extraction for their pattern recognition. In general, large data decreases prediction accuracy, and also not all features are crucial to detecting the label of the data class [34]. This section gives information about the two datasets used in this research, feature selection algorithms used in layer 1 of the model, various classification models, and hyper parameter optimization used in layer 2, for the performance enhancement of the model.

### A. DATASETS
#### 1) DESCRIPTION OF KAGGLE'S HEART DISEASE DATASET
This cardiac dataset is obtained from Kaggle and is made up of 70, 000 patient records with a total of 12 features. There is an in-depth explanation of each feature and corresponding value in Table 2. A person's risk for heart disease is determined by these attributes. Three types of features were identified in this dataset such as objective, examination, and subjective.
- The objective feature type indicates the information related to the patient, like age, height, weight, and gender.

predominantly focused on small datasets in cardiovascular disease prediction using machine learning techniques. To overcome this limitation, the study utilizes two large

**TABLE 2.** Description of Dataset-1.

| Feature (Type) | Description | Value |
|---|---|---|
| Age (Objective) | Age | In days |
| Height (Objective) | Height | In cms |
| Weight (Objective) | Weight | In kgs |
| Gender (Objective) | Men/Women | Women(1), men(1) |
| Ap-hi (Examination) | Systolic BP | Integer values |
| Ap-lo (Examination) | Diastolic BP | Integer values |
| Cholesterol (Examination) | Cholesterol level | Level 1,2,3 |
| Gluc (Examination) | Glucose level | Level 1,2,3 |
| Smoke (Subjective) | Smoking habit | Binary value |
| Aloc (Subjective) | Alcohol intake | Binary value |
| Active (Subjective) | Physical activity | Binary value |
| Cardio (Target variable) | Disease yes/no | 1-disease, 0-normal |

**TABLE 3.** Description of Dataset-2.

| Feature & Description | Value |
|---|---|
| Age | In years |
| Sex | Female (0), Male (1) |
| Cp (Type of chest pain) | 4 values (0,1,2,3) |
| Restbps (Resting bp value ) | In mm Hg (94-200) |
| Chol (Serum cholesterol) | In mg/dl (126-564) |
| Fbs (Fasting blood sugar) | True (1), False (0) |
| Restecg (Resting ECG results) | (0-2) |
| Thalach (Maximum heart rate) | (71-202) |
| Exang (Exercise induced angina) | Yes (1), no (0) |
| Oldpeak (ST depression) | (0-6.2) |
| Slope (Slope of peak ST segment) | (0-2) |
| Ca (Count of major vessels) | (0-3) |
| Thal (thalassemia) | (1,2,3) |
| Target (Disease present or not) | Disease (1), normal (0) |

- The examination feature type comprises patient data obtained from the outcomes of a medical examination.
- The subjective feature type includes information provided by the patient concerning their habits and lifestyle.

### 2) DESCRIPTION OF UCI'S HEART DISEASE DATASET

This cardiac dataset is downloaded from Kaggle and has 1025 patient records with 14 attributes. There are total of 312 female records and 713 male records in the dataset, and there are 499 normal cases and 526 patient cases. Table 3. contains detailed information about all features.

### B. FEATURE SELECTION METHODS
### 1) ANOVA F-TEST

Analysis of variance (ANOVA) is a statistical test used to assess the ratio between two variances. The scikit-learn library's f classif() method is employed to generate the ANOVA F-scores between every feature and the target feature. The formula for ANOVA F-test is represented as

$$F = \frac{Varience\ between\ the\ groups}{Varience\ within\ the\ groups} \quad (1)$$

$$varience\ between\ the\ groups = \frac{\sum_{i=1}^{n} n_i (\bar{Y}_i - \bar{Y})^2}{(K-1)} \quad (2)$$

$$varience\ with\ in\ the\ groups = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{(N-K)} \quad (3)$$

where Y is the overall mean of the data, N is the overall sample size, K is the number of groups, $Y_i$ is the $i^{th}$ group sample mean, $n_i$ is the number of observations in that group, and $Y_{ij}$ is the observation in the $i^{th}$ out of K groups [35].

### 2) MUTUAL INFORMATION

Mutual information (MI) determines whether two random variables are mutually dependent. The entropy of a random variable and mutual information are closely related to each other. Mutual information scores are calculated by using the mutual_info_classif () function to find the relationship between the features and the target feature. MI can be calculated in terms of entropy H(X) and conditional entropy H(X/Y) is represented as.

$$MI\ (X{:}Y) = H\ (X) - H\left(\frac{X}{Y}\right) \quad (4)$$

Where the H(X) entropy for the values {x1, x2, x3, x4,..., xn} can be written as

$$H\ (X) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \quad (5)$$

Conditional entropy $H\left(\frac{X}{Y}\right)$ can be written as

$$H\left(\frac{X}{Y}\right) = -\sum_{i=1}^{n} \sum_{j=1}^{n} P\ (x_i, y_i)\ log_2 \frac{P\ (x_i)}{P\ (x_i, y_i)} \quad (6)$$

where $P\ (x_i, y_i)$ is X = $x_i$ and Y = $y_i$ joint probability, $P(x_i)$ is the probability mass function of $x_i$ [36].

### 3) CHI-SQUARE TEST

For feature selection, the Chi-square test uses the univariate statistical approach, which identifies the correlation between the features. Chi-square scores are calculated using the chi2() function from the scikit-learn machine library to determine the independence of two features. Using SelectKBest (), the most effective features with the highest Chi-square scores were selected [37].

$$X^2 = \sum \frac{(f_0 - f_E)^2}{f_E} \quad (7)$$

Here observer frequency is denoted as $f_0$ and expected frequency is denoted as $f_E$.

### C. CLASSIFICATION ALGORITHMS
### 1) DECISION TREE

A decision tree is a tree-structured classifier whose evaluation is like a graphical representation for discovering all possible solutions to a given problem. There are usually several levels of nodes in a decision tree, with the topmost node known as the root node, and the rest known as children. Each internal node represents the evaluation of an input variable or feature.

Upon evaluation, the classification techniques branch to the appropriate child node, where the evaluation and branching process continues. Using high entropy inputs, trees are constructed for training samples of data D. In this simple and fast trees are constructed using a top-down recursive divide and conquer strategy. As part of the tree-pruning process, irrelevant samples are removed from D [18].

$$Entropy = -\sum_{j=1}^{m} p_{ij} \, log_2 \, p_{ij} \qquad (8)$$

#### 2) RANDOM FOREST

The Random Forest is an ensemble learning technique that merges multiple decision trees using the bagging concept to improve prediction accuracy. To build a committee, this method employs bagging and random subspace sampling. As a result, each data instance's final class label is selected by a majority vote.

Let {X, T} denote a set of training data where X = $x_0$, $x_1$, $x_2$ …, $x_{n-1}$ and T = $t_0$, $t_1$, $t_2$ …, $t_{n-1}$. If h(x) represents a classification tree, the method selects random samples with the replacement of previously collected training data on h(x) and uses these samples to train h(x). In this procedure, model variance is reduced without causing model bias to increase, resulting in improved model performance [17].

#### 3) SUPPORT VECTOR MACHINE

The Support Vector Classifier (SVC) operates by identifying the hyperplane that effectively segregates the classes while maximizing the gap between them. Additionally, SVC can handle non-linear data by using techniques such as kernel tricks to transform the data into higher dimensions, allowing for more complex decision boundaries. The SVC algorithm is well-suited for managing both small and large datasets that are otherwise challenging to handle. It leverages different kernel types, such as linear, radial basis function (RBF), polynomial, and sigmoid, to create an effective prediction model. RBF kernel is used in the proposed model [26].

#### 4) MULTI-LAYER PERCEPTRON

A multilayer perceptron (MLP) contains many layers, as its name suggests. Perceptron with only one layer solves linearly separable problems, whereas MLP solves nonlinearly separable problems by adding one or more layers to a single layer. The most common uses of these algorithms are pattern recognition, classification of input patterns, prediction depending on input information, and approximation. MLPs are feedforward neural networks with at least three layers: input, hidden, and output. Neurons use activation functions to communicate with one another. For training, it uses a back-propagation supervised learning technique; it can distinguish new data [22].

#### 5) HYPER PARAMETER TUNNING WITH GRID SEARCHCV

Machine learning applications train many models on data, and then select the best model based on its performance. To make progress, we cannot say with certainty that a certain

**TABLE 4.** Confusion matrix.

| Feature | Description |
|---------|-------------|
| True +ve | classifier predicted +ve, original outcome is true (TP) |
| True -ve | classifier predicted -ve, original outcome is true (TN) |
| False +ve | classifier predicted +ve, original outcome is false (FP) |
| False -ve | classifier predicted -ve, and original  is false (FN) |

model is the best fit. Consequently, we strive to improve the model in any way we can. A key characteristic of models' performance is that when the hyperparameters are selected correctly, it can significantly enhance the model's output. The Grid Search method is widely used for determining all hyperparameter combinations. Two of the most important parameters in Grid Search are the learning rate and the number of layers. Initially, each hyperparameter is assigned a set of values. During each cycle, the combination of hyperparameters is determined. At the end of the learning process, the most successful combination of hyperparameters is selected and implemented.

#### D. EVALUATION METRICS

Various measures are applied to estimate the accuracy of the model, with the majority of these being derived from the values within the confusion matrix. The confusion matrix is comprised of four key values, as elaborated in Table 4.

In the validation of ML algorithms, four statistical metrics have been employed, including accuracy, precision, F1-score, and recall [38]. Accuracy quantifies the number of correct predictions within both the positive and negative classes. Precision measures the proportion of true positives among the predicted positive cases. Recall signifies the ratio of correctly predicted positive instances to all actual positive cases. Precision and recall are both considered in computing the F1 score.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100 \qquad (9)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \qquad (10)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \qquad (11)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \qquad (12)$$

#### E. TLV METHODOLOGY

A TLV (Two-Layer Voting) model is designed for the diagnosis of cardiovascular issues in two scenarios with small and large datasets. The proposed method performance is examined using UCI's heart disease dataset and heart disease datasets collected from Kaggle. Algorithm 1 explains the two-layered architecture in detail. Here's a detailed description of the TLV method:

**Layer 1 - Feature Selection and Voting:** In the first layer, the TLV method performs feature selection using hard voting of multiple statistical methods, such as ANOVA f-test, Chi-squared test, and Mutual Information. Each statistical

method independently evaluates the significance of features in the dataset and assigns scores or ranks to them based on their relevance to the classification task. Hard voting is then applied to select the most significant features by considering the votes of all three statistical methods. Features that receive the most votes are retained for further processing.

**Layer 2 - Classification and Voting:** In the second layer, the TLV method applies multiple base classifiers, such as Random Forest, Decision Tree, Support Vector Classifier, and Multi-Layer Perceptron, to the selected features. Each base classifier independently generates predictions for the target variable based on the input features. Soft voting is employed in this layer, where each base classifier assigns probabilities or confidence scores to its predictions instead of binary outcomes. The final classification decision is made by combining the probability scores from all base classifiers using a weighted average or another aggregation method.

**Hyperparameter Tuning:** To optimize the performance of the TLV method, hyperparameter tuning is conducted for each base classifier using techniques like GridSearchCV. Hyperparameter tuning involves systematically searching for the best combination of model parameters that maximizes the performance metrics, such as accuracy, precision, recall, or F1-score. By fine-tuning the parameters of each base classifier, the TLV method can adapt to the specific characteristics of the dataset and improve its predictive capabilities.

**Evaluation and Result Analysis:** After the TLV method has been applied to the dataset, its performance is evaluated using appropriate evaluation metrics, such as accuracy, precision, recall, or area under the ROC curve (AUC). The results obtained from the TLV method are compared against those of other classification techniques or baseline models to assess its effectiveness and superiority. Additionally, the contribution of each base classifier and feature selection method to the overall performance of the TLV method is analyzed to gain insights into its functioning and potential areas of improvement. In summary, the Two-Layered Voting method combines feature selection and classification through a hierarchical approach, leveraging the collective intelligence of multiple classifiers to enhance the accuracy and reliability of classification models in machine learning tasks.

A workflow of the proposed TLV model is illustrated in Fig. 1.

- The heart disease patient's data is gathered from the two datasets, cleaned and preprocessed, including handling missing data, and label grouping.
- Based on the feature analysis new features Pulse Pressure (PP), Body Mass Index (BMI), and Mean Arterial Pressure (MAP) are generated to improve the model's accuracy.
- K-modes are applied to find the optimum point in the data
- In the first layer, three statistical methods are used to select the most prominent features, namely the ANOVA f-test, the Chi-squared test, and the Mutual Information test.

---

**Algorithm 1** Two Layered Voting Framework

**Input:**
H= Heart disease dataset
$f_i = \{f_1, f_2,\ldots, f_n\}$ \\ features
$L_i = \{L_1, L_2, L_3\}$ \\ Set of classifiers
$FS_i = \{FS_1, FS_2, FS_3\}$ \\ Set of feature selection algorithms
$D = \{(x_i, y_i) \mid i=1, 2,\ldots, n\}$ \\ Training set
$T = \{(xt_i) \mid i=1, 2,\ldots, n\}$ \\ Testing set
**Output:**
Predict class labels of the test samples $O = \{(yt_i) \mid i=1,2,\ldots,n\}$
**Feature selection-LAYER: 1**
**Procedure** FEATURE_SELECTION $(f_1, f_2,\ldots, f_n)$
   $FS_1$ = Anovaf_Test$(f_1, f_2,\ldots, f_n)$
   $FS_2$ = Mutual_Information$(f_1, f_2,\ldots, f_n)$
   $FS_3$ = Chi_squared$(f_1, f_2,\ldots, f_n)$
   Hvote_1=hard voting of $(FS_1, FS_2, FS_3)$
   Svote_1=soft voting of $(FS_1, FS_2, FS_3)$
   $f_s$ = top_features from the votes of (Hvote_1, Svote_1)
   **return** $f_s$
**end procedure**
**Classification LAYER: 2**
**Procedure** CLASSIFICATION $(f_s, L_i)$
   Training_data, Testing_data = split(attributes,label)
   $L_1$ = Classifier_1 with GridSearchCV$(f_s)$
   $L_2$ = Classifier_2 with GridSearchCV$(f_s)$
   $L_3$ = Classifier_3 with GridSearchCV$(f_s)$
   Hvote_2=hard voting of $(L_1, L_2, L_3)$
   Svote_2=soft voting of $(L_1, L_2, L_3)$
   Predictions= Best value (Hvote_2, Svote_2)
   Return $O = \{yt_i\}$
**end procedure**

---

- The final results are derived by examining the outcomes of both soft and hard voting across the three models. Following data preprocessing and feature selection in layer 1, the dataset is then split into training and testing sets.
- Models such as Decision Tree, Multi-Layer Perceptron, Random Forest, and the SVC are employed in the second layer for classification. The final results in this layer are obtained by comparing the outcomes of soft and hard voting.
- Hyperparameter tuning is applied using GridsearchCV to all the classifiers in the second layer.
- The TLV model is tested using various metrics namely accuracy, precision, recall, and F1-score, and benchmarked with the existing models.
- The proposed model works well for the binary classification of coronary artery disease.

## IV. RESULTS AND DISCUSSION
### A. EXPERIMENTAL RESULTS WITH DATASET-1
Google Colab was used to implement all algorithms, which were worked on data from Kaggle's cardiovascular disease dataset with 70,000 instances.
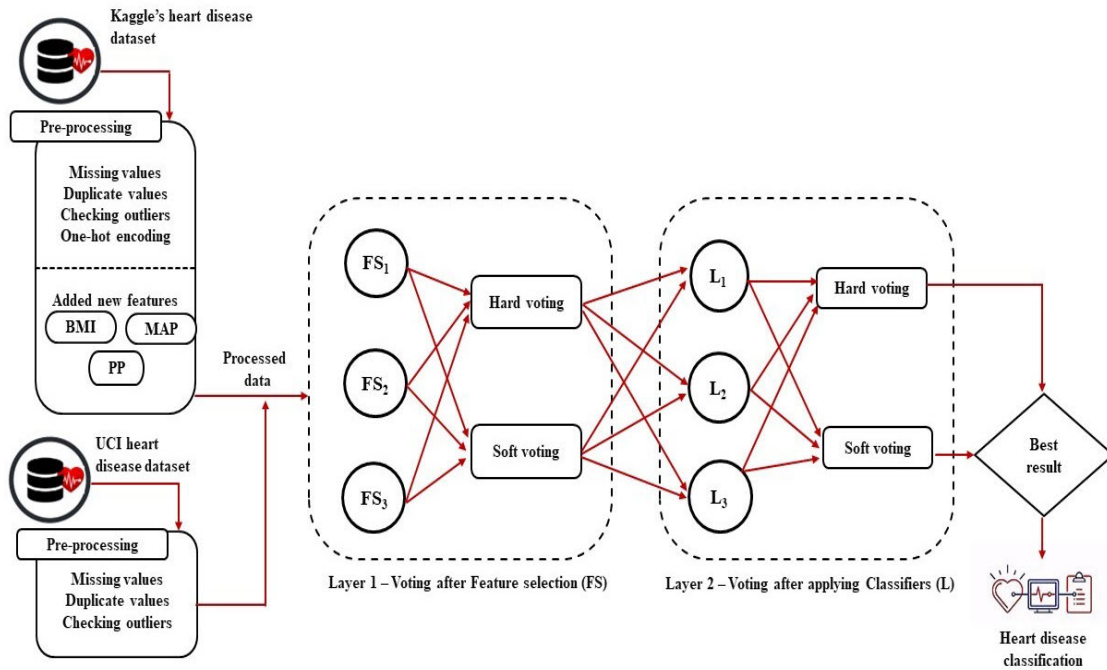
**FIGURE 1.** Basic workflow of proposed TLV (Two-Layer Voting) model.

### 1) FEATURE CREATION

By implementing feature engineering, we have generated additional features that enhance the accuracy of our models. Notably, the development of cardiac issues is closely associated with BMI. BMI serves as a reliable indicator of an individual's body fat percentage and susceptibility to associated health conditions. Research indicates that individuals with higher BMI values are at an increased possibility of acquiring specific diseases, particularly cardiac issues [39].

Incorporated into our existing features, we have included BMI, derived from the individual's height and weight, as a new feature. Additionally, MAP stands as another vital feature that we have introduced to this dataset. MAP measurement provides valuable insights into blood flow, resistance, and arterial pressure. Notably, it serves as a significant measure in various disease conditions, including cardiac diseases. MAP values can be created by using the systolic and diastolic BP readings [40]. The measurement of PP may indicate the likelihood of developing heart disease. A person's pulse pressure is currently an indication of their risk of coronary heart disease, especially when they are middle-aged or older [41]. Formulas for BMI, MAP, and PP are written as below equations.

$$BMI = weight / height^2 \qquad (13)$$

$$MAP = \frac{systolic\ pressure + (2 \times diastolic\ pressure)}{3} \qquad (14)$$

$$PP = systolic - diastolic \qquad (15)$$

### 2) K-MODES CLUSTERING

By clustering, a group of instances is combined according to similarity measures. The k-modes algorithm is the same as the k-means algorithm, except that it takes dissimilarity measures for categorical data and replaces cluster means with modes. As a result, categorical data can be handled effectively by the algorithm.
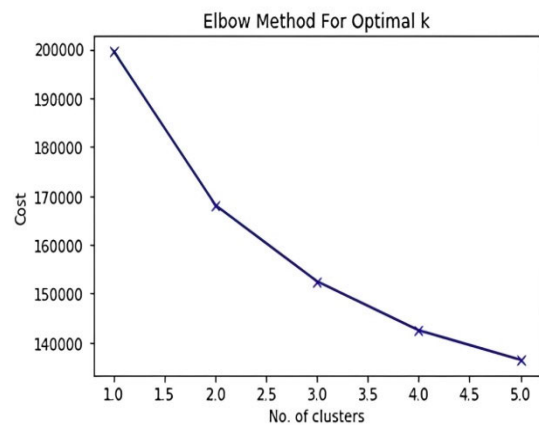


**FIGURE 2.** Elbow method for all data.

The elbow curve with Huang initialization is used first to find the optimal number of clusters. This generates a k-mode model with those clusters, fits the model to the data, and then determines the cost, which is the distance between each cluster's attribute modes and the data points.

To determine the optimal cluster size, the costs are plotted on an elbow graph. In the elbow method, the plot of costs is examined for a "knee" or inflection point, which signifies the point where adding more clusters doesn't make a significant improvement in model fit. Women and men have different biological characteristics that can affect the manifestation and
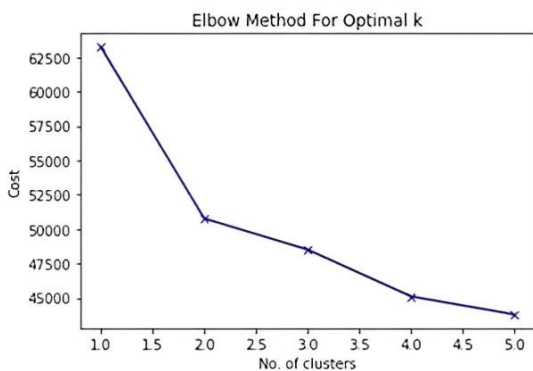
progression of diseases, which may make it advantageous to split the dataset by gender for prediction. A man's risk of developing heart disease may differ from that of a woman, and their symptoms may be different as well.

A separate analysis of men's and women's data can identify the most important risk factors and patterns of disease progression that are not clear upon consolidation of the data. Following that, the elbow curve method was applied to identify the optimal number of clusters for both males and females. According to Fig.2 and Fig.3, for both male and female datasets, the knee point was located at 2.0, indicating that 2 was the optimal number of clusters. Moreover, the correlation between different categories is determined by a correlation table. From Fig.4, gender, PP, glucose, MAP, BMI, and smoke are highly correlated factors. With the help of this matrix, you can also determine the intra-feature dependency.



**FIGURE 4.** Correlation values of all features.



**(a)**

and soft voting classifiers have been formed in the first layer with three feature ranking algorithms namely the ANOVA f-test, the Chi-squared test, and the Mutual Information. The rankings from hard and soft voting feature extraction methods are represented by using a heat map in Fig.5.
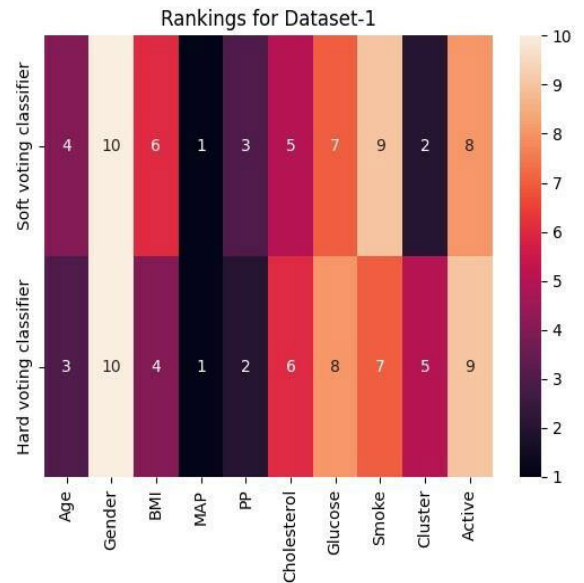


**FIGURE 5.** Ranking of feature selection methods-(dataset-1).



**(b)**

**FIGURE 3.** Elbow method for (a) male and (b) female.

### 3) RESULTS WITH TLV
In the first layer of the TLV model feature extraction is done by hard and soft voting. A total of 15 attributes are there in the dataset.

One form of meta-classifier is the voting algorithm, which assembles variant ML algorithms for prediction. The hard

After collecting the rankings of each algorithm, we applied both hard and soft voting for feature extraction. The top 10 features from both voting approaches are extracted and recorded as two separate subsets. As depicted in Fig.5, while there is a slight variance in the rankings between soft and hard voting, the top 10 features remain largely consistent. Consequently, both feature subsets comprise the same features, such as cluster, gender, age, BMI, MAP, cholesterol, glucose, smoke, active, and PP. These 10 features are extracted from

layer 1 for further investigation. Furthermore, the dataset is partitioned into train as 75, and test as 25 ratios.

The soft and hard voting classifier has been formed in the second layer with Random Forest, Decision Tree, and MLP. GridSearchCV is applied for hyper-parameter tuning. After the model has completed training, 5-fold cross-validation is applied to classify the risk of cardiovascular problems such as CAD. All three ML classifiers' confusion matrix values on testing data are represented in Fig.6, Fig.7 and Fig. 8. We have evaluated test and train accuracies for the ML classifiers as represented in Fig.9. The soft and hard voting classification models utilize three ML algorithms with GridSearchCV.
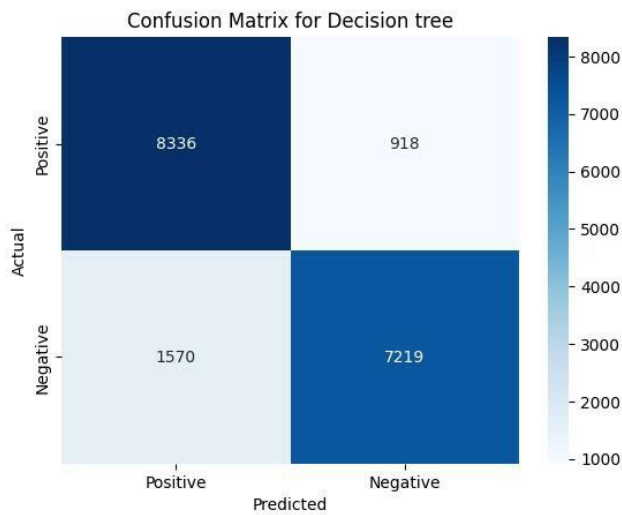


**FIGURE 6.** Confusion matrix values for DT on dataset-1.
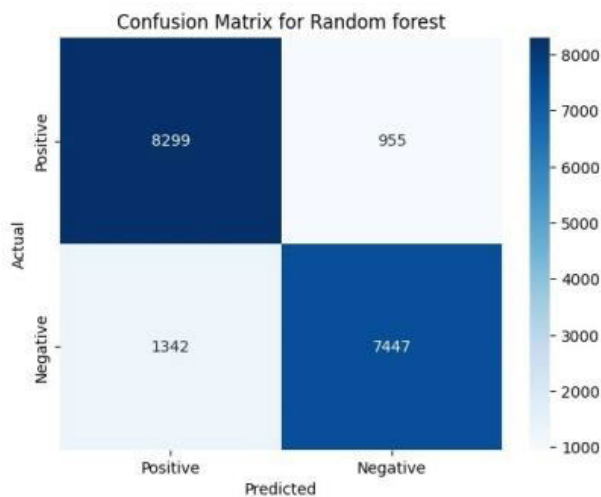


**FIGURE 7.** Confusion matrix values for RF on dataset-1.

As explained in Table 5, the MLP classifier exhibited the smallest disparity, with a difference value of 0.4 between test and train accuracies. In contrast, the decision tree registered the highest difference value at 2.8. Employing these three
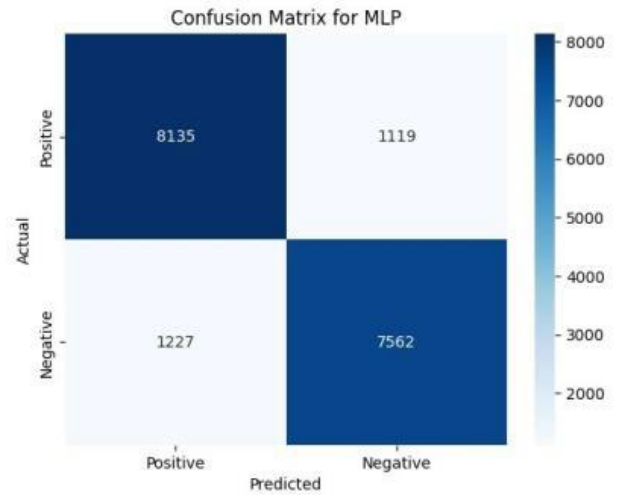


**FIGURE 8.** Confusion matrix values for MLP on dataset-1.
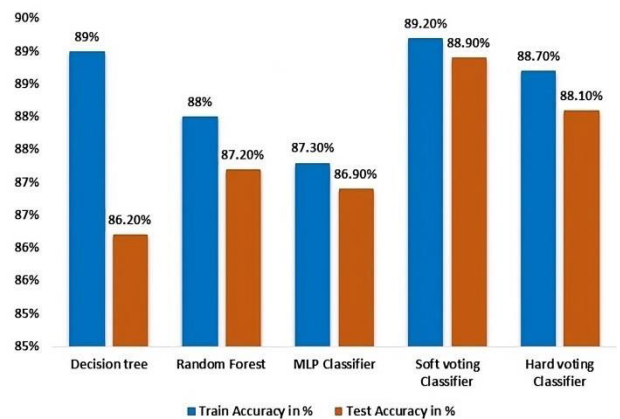


**FIGURE 9.** Train and test accuracy analysis for dataset 1.

classifiers, the proposed TLV model assessed two scenarios through soft and hard voting. The findings indicate that the proposed TLV model, when employing soft voting, yielded a minimal difference of 0.3.

As shown in Table 5, among the tested classifiers, Random Forest exhibited the highest accuracy at 87.2%, MLP demonstrated the highest precision at 86.8%, and Decision Tree displayed the highest recall at 90%. Soft and hard voting classifiers were developed based on these three classifiers, with Random Forest assigned the highest weight due to its superior performance in the soft voting scenario. The proposed model was evaluated under both soft and hard voting scenarios, resulting in 88.9% accuracy for soft voting and 88.1% for hard voting, as illustrated in Fig. 10.

Performance evaluation results for the soft voting model in layer 2 are depicted in Fig. 11.

## B. EXPERIMENTAL RESULTS WITH DATASET-2
1) The second dataset was obtained from Kaggle and consists of four merged databases with 1025 records, with 713 males and 312 females. There are 526 abnormal heart patients in this
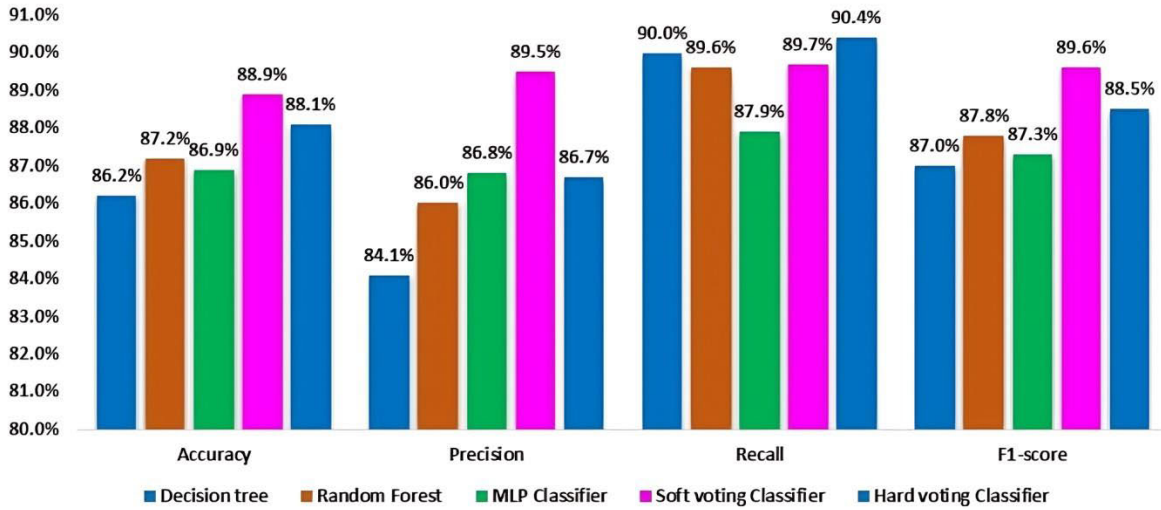
**FIGURE 10.** Performance analysis of proposed TLV (Two-Layer Voting) model with soft and hard voting for dataset-1.

**TABLE 5.** (a) Test and train accuracy difference for Dataset 1.
(b) Performance analysis of proposed TLV model for Dataset 1.

|  (a)  |  |  |  |
|---|---|---|---|
| Model | Train accuracy | Test accuracy | Difference |
| Decision tree | 89 | 86.2 | 2.8 |
| Random forest | 88 | 87.2 | 0.8 |
| MLP classifier | 87.3 | 86.9 | 0.4 |
| TLV with soft voting | 89.2% | 88.9% | 0.3 |
| TLV with hard voting | 88.7% | 88.1% | 0.6 |

|  (b)  |  |  |  |  |
|---|---|---|---|---|
| Model | Accuracy | Precision | Recall | F1-score |
| Decision tree | 86.2% | 84.1% | 90% | 87% |
| Random forest | 87.2% | 86% | 89.6% | 87.8% |
| MLP classifier | 86.9% | 86.8% | 87.9% | 87.3% |
| TLV with soft voting | 88.9% | 89.5% | 89.7% | 89.6% |
| TLV with hard voting | 88.1% | 86.7% | 90.4% | 88.5% |



**FIGURE 11.** Accuracy analysis for dataset-1.

dataset and 499 healthy people. Therefore, the percentage of Class 1 is approximately 51.32%, and the percentage of Class 0 is approximately 48.68% out of the total of 1025. During the pre-processing stage, a process of checking for missing, and duplicate values was conducted on the dataset, and no instances of missing values were identified. Afterward, the dataset is separated as 80% data for the train and 20% for test purposes.

### 1) FEATURE SELECTION AND RESULTS WITH TLV
In the first layer of the TLV model feature extraction is done by hard and soft voting. As previously mentioned, three
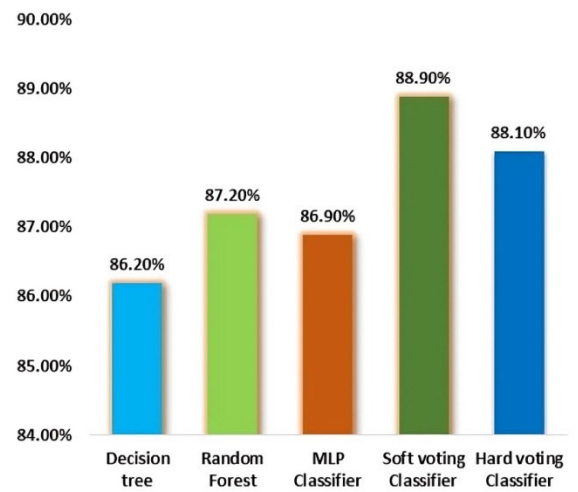
algorithms are applied in the first layer namely the ANOVA f-test, the Chi-squared test, and the Mutual Information. The rankings from the soft and hard feature selection methods are represented by using a heat map in Fig.12 to dataset 2.

Comparisons between the top 10 features from the soft and hard voting layers reveal a shared set, including old peak, thalach, cp, exang, ca, chol, thal, age, slope, and sex. In the subsequent layer, classification is performed using Random Forest, Decision Tree, and SVC. GridsearchCV is applied for hyper-parameter tuning for the classification of the CAD. For dataset 2 all three ML classifiers' confusion matrix values on testing data are represented in Fig.13, Fig.14, and Fig.15.

Table 6 presents the disparities in train and test accuracies, along with comparisons of soft and hard voting performances. The Random Forest classifier exhibits the smallest difference value at 0.3 between test and train accuracies, while the MLP classifier demonstrates the highest difference at 2.6.
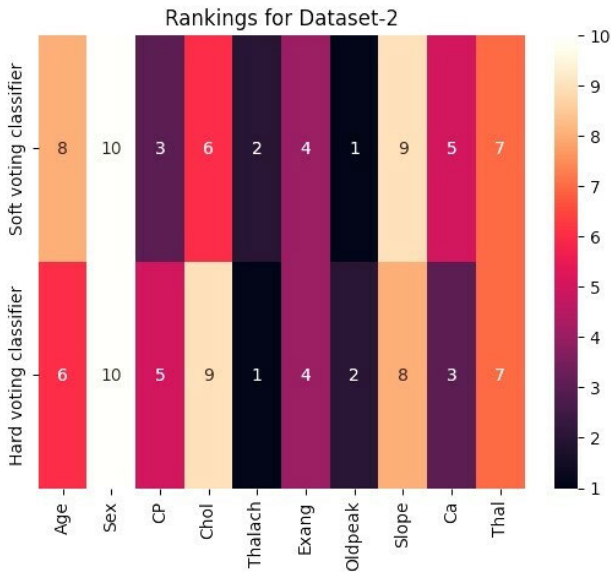
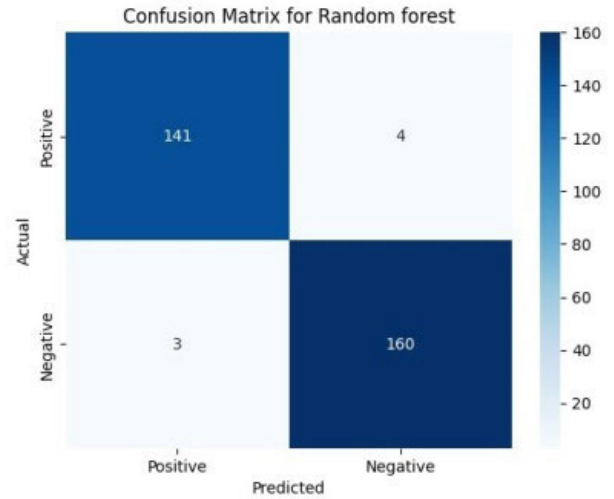**FIGURE 12.** Ranking of feature selection methods for dataset-2.



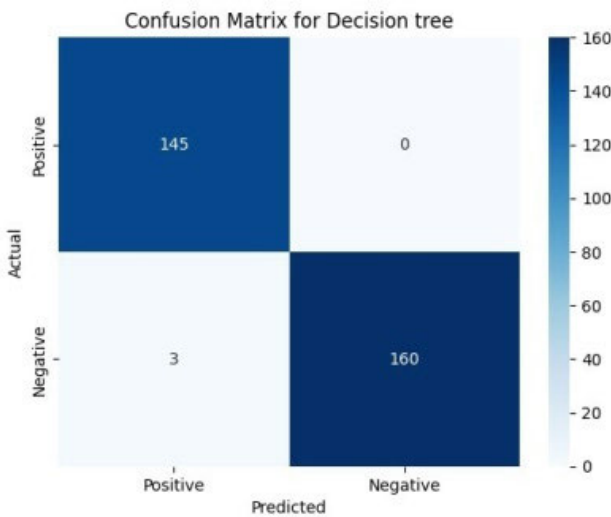**FIGURE 13.** Confusion matrix values for DT for dataset-2.


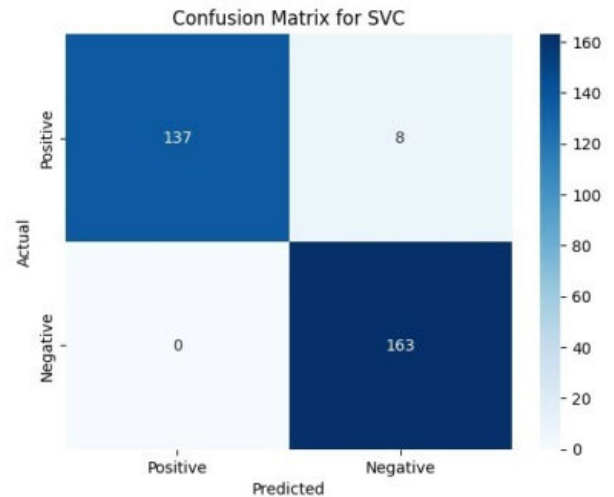
**FIGURE 14.** Confusion matrix values for RF for dataset-2.



**FIGURE 15.** Confusion matrix values for SVC for dataset-2.

**TABLE 6.** Test and train accuracy difference for Dataset 2.

| Model | Train accuracy | Test accuracy | Difference |
|---|---|---|---|
| Decision tree | 100 | 99 | 1.0 |
| Random forest | 98 | 97.7 | 0.3 |
| MLP classifier | 100 | 97.4 | 2.6 |
| TLV with soft voting | 99.4 | 99.3 | 0.1 |
| TLV with hard voting | 99 | 98.8 | 0.2 |

By employing three classifiers, we conducted a comprehensive comparison of soft and hard voting results within the proposed TLV model. Notably, the proposed TLV model with soft voting shows a minimum difference value of 0.1, and the TLV model with hard voting has shown a difference value of 0.2. Consequently, we infer that in our proposed model, soft voting surpasses all algorithms, achieving the highest accuracy.

As shown in Table 7 three ML algorithms such as Decision Tree, Random Forest, and SVC with GridsearchCV are used in the soft and hard voting classification models.

Test and train accuracies were evaluated for the ML classifiers, as depicted in Fig.16. Notably, the MLP classifier demonstrated the largest difference between training and testing accuracies, while the proposed model with the soft voting classifier showcased a minimal difference.

The soft voting classification model attained an exceptional accuracy of 99.3%. Among the three classifiers assessed, Decision Tree exhibited the highest accuracy of
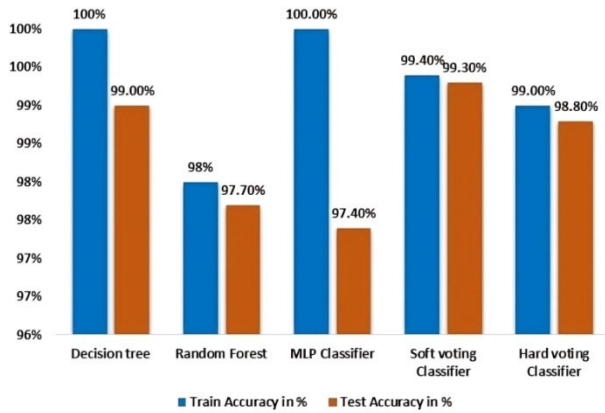
**FIGURE 16.** Test and train accuracies of all algorithms.

**TABLE 7.** Performance analysis of proposed TLV model for Dataset 2.

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Decision tree | 99 | 97.9 | 100 | 98.9 |
| Random forest | 97.7 | 97.9 | 97.2 | 97.5 |
| MLP classifier | 97.4 | 100 | 94.4 | 97.1 |
| TLV with soft voting | 99.3 | 98.6 | 100 | 99.3 |
| TLV with hard voting | 99 | 99.3 | 98.6 | 98.9 |

99%, SVC displayed the highest precision of 100%, and Decision Tree showcased the highest recall of 100%.

Given the superior results of the Decision Tree, it was assigned a greater weight in the soft vote classifier. The TLV model, employing soft voting, achieved impressive outcomes with a 99.3% accuracy, 98.6% precision, 100% recall, and 99.3% f1-score. While the hard voting layer attained a 99% accuracy, slightly lower than the soft voting results. The performance evaluation results for the proposed TLV voting model in layer 2 are illustrated in Fig.17 and Fig.18.

## C. COMPARISON OF PROPOSED TLV PREDICTION MODEL WITH LITERATURE SURVEY STUDIES AND FUTURE WOK

To conclude, we compared and analyzed the experimental findings of the TLV model with those of early studies that are presented in the literature survey; the results are detailed in Table 8 and Table 9. These findings conclude that the TLV model outshined all earlier studies on the severity level classification using two Kaggle's heart disease datasets. The soft voting prediction relies on the arg max of the sum of the predicted probabilities. Assigning weights to the top-performing classifiers in soft voting enables more influential models to have a greater impact on the final prediction, thereby improving the ensemble model's robustness and accuracy. In our approach, we allocated more weight to the most effective classifier, resulting in favorable outcomes compared to the findings mentioned in the related work.
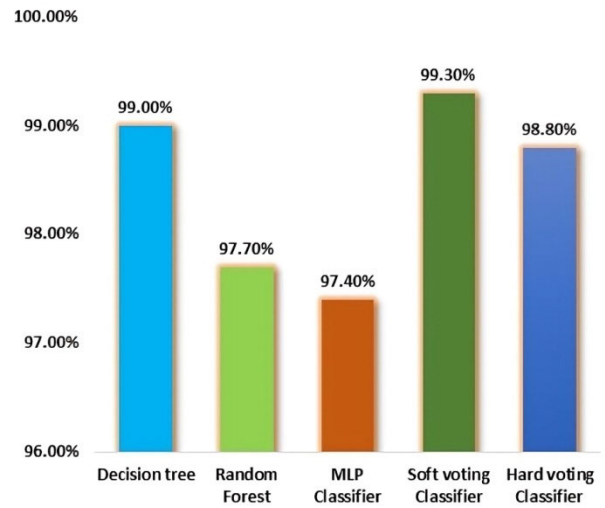


**FIGURE 17.** Accuracy analysis for dataset-2.

**TABLE 8.** Performance analysis of TLV model compared with previous studies on UCI's dataset with 1025 patient records.

| Methodology | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Gradient descent [12] | 98.54 | - | 99.4 | - |
| Linear SVC & DNN [13] | 98.56 | 97.8 | 99.3 | 98.3 |
| Bagging ensemble [14] | 98.6 | 97.8 | 99 | 98.5 |
| XGBoost [15] | 95.61 | - | - | - |
| J48 [16] | 98.4 | - | - | - |
| Random forest [17] | 98.0 | - | - | - |
| Decision tree [18] | 98.8 | - | - | - |
| Proposed TLV model with soft voting classifier | 99.3 | 98.6 | 100 | 99.3 |

Many authors only analyzed small-sized datasets in their study but didn't achieve a high accuracy to a large-sized dataset since large and diverse data affected the working of the ML model. The best machine learning hyperparameters vary according to the task and dataset being used. The GridsearchCV algorithm enhances ML performance by discovering the best hyperparameter settings for balancing techniques and classifiers.

The proposed model demonstrated the highest accuracy compared with the literature survey studies, there is a need for improvement in its performance when applied to Kaggle's dataset. As both datasets employed in this experiment are balanced, we did not specifically focus on addressing class imbalance issues in the proposed model. Furthermore, the proposed model requires a detailed analysis of its time complexity, as this aspect has not been thoroughly addressed thus far in this paper. Additionally, we recognize the importance of updating the datasets to align with present scenarios. It is
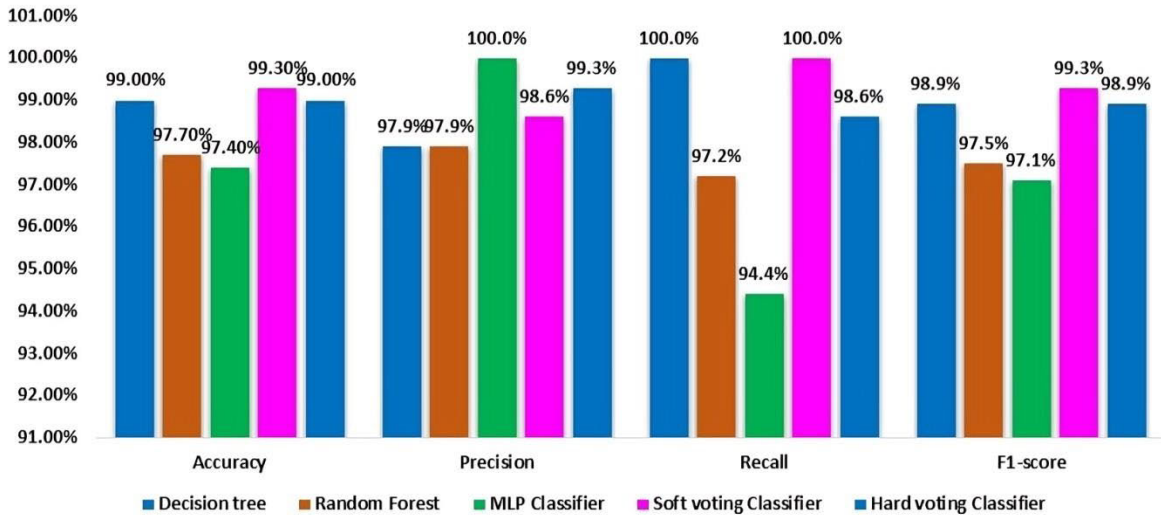
**FIGURE 18.** Performance analysis of proposed TLV (Two-Layer Voting) model for dataset-2.

**TABLE 9.** Performance analysis of TLV model compared with previous studies on Kaggle's dataset with 70000 patient records.

| Methodology | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Random forest [25] | 73 | 65 | 80 | - |
| XGBoost & wrapper methods [26] | 73.74 | 76 | 69 | 72 |
| Bagged decision tree [27] | 74.8 | 76.2 | 67.4 | 71.5 |
| Optimized decision tree [29] | 73.54 | 75.82 | 68.89 | 78.16 |
| Neural networks [30] | 71.82 | 72 | 72 | 72 |
| Random forest [31] | 74 | - | - | - |
| Random forest [32] | 69.6 | 72.1 | 65.49 | - |
| MLP classifier [33] | 87.28 | 88.7 | 84.8 | 86.7 |
| Proposed TLV model with soft voting classifier | 88.9% | 89.5% | 89.7% | 89.6% |

essential to validate the model's predictions by comparing them with actual laboratory data to ensure their reliability. In our present research, we recognize the opportunity for improvement in incorporating datasets with more diverse information from various hospital facilities across different regions. Addressing this gap in future work will empower us to identify patterns in the data at a national level and provide robust support for the model's generalization.

In our future research, we will emphasize the real-time analysis of cardiovascular datasets, including a comprehensive evaluation of the time complexity of the implemented algorithms. The incorporation of real-time datasets in cardiac disease classification is crucial for ensuring the accuracy and relevance of predictive models, enabling dynamic insights into the evolving patterns of cardiovascular health. By utilizing up-to-date data, the models can adapt to current trends,

enhancing prediction precision and enabling timely interventions for at-risk individuals, thus improving overall patient outcomes. The changes in lifestyle, dietary patterns, and post-COVID conditions may introduce a range of new risk factors contributing to the development of cardiovascular conditions. By addressing these aspects, we aim to contribute significantly to the advancement of cardiovascular disease detection and prognosis.

## V. CONCLUSION
In this experimental study, an effective and accurate disease classification system TLV (Two-Layer Voting) is proposed to find the absence or presence of coronary artery disease (CAD). The proposed model integrates hard voting and soft voting with parameter optimization methods for classification. Unlike the other studies, this study has taken two different-sized datasets Kaggle's heart disease dataset of over 70,000 records and UCI's heart disease dataset of 1025 records. By the experimental results, it has been proved that in layer 1 the hard voting framework with the ANOVA f-test, Chi-squared test, and Mutual Information is very effective for feature selection for any dataset regardless of its size. The experimental results show that soft voting in the second layer with the MLP algorithm, the Decision Tree, the SVC, and the Random Forest algorithms were more effective in achieving higher-quality results. Optimizing hyperparameters using GridSearchCV helps return the best model fit by determining the best-tuned parameters. A comparison is made between the proposed model and previous studies. An analysis of our TLV model's work has been presented by applying various performance measures. The proposed TLV got an accuracy of 99.3% on UCI's heart disease dataset and 88.9% on Kaggle's heart disease dataset. These findings revealed that our proposed TLV methodology with soft voting accomplished accuracy, precision, recall, and f1-score of up to 99.3%, 98.6%, 100%, and 99.3% for the UCI's

heart disease dataset, 88.9%, 89.5%, 89.7%, and 89.6% for Kaggle's heart disease dataset respectively. These findings may be useful to healthcare professionals in predicting CAD and improving patient care.

## REFERENCES

[1] L. R. Teixeira, F. Pega, A. M. Dzhambov, A. Bortkiewicz, D. T. C. da Silva, C. A. F. de Andrade, E. Gadzicka, K. Hadkhale, S. Iavicoli, M. S. Martínez-Silveira, M. Pawlaczyk-Łuszczyńska, B. M. Rondinone, J. Siedlecka, A. Valenti, and D. Gagliardi, "The effect of occupational exposure to noise on ischaemic heart disease, stroke and hypertension: A systematic review and meta-analysis from the WHO/ILO joint estimates of the work-related burden of disease and injury," *Environ. Int.*, vol. 154, Sep. 2021, Art. no. 106387, doi: 10.1016/j.envint.2021.106387.

[2] C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han, and J. Yu, "Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the Internet of Medical Things platform," *IEEE Access*, vol. 8, pp. 59247–59256, 2020, doi: 10.1109/ACCESS.2020.2981159.

[3] M. Kivimäki and I. Kawachi, "Work stress as a risk factor for cardiovascular disease," *Current Cardiol. Rep.*, vol. 17, no. 9, pp. 1–9, Aug. 2015, doi: 10.1007/s11886-015-0630-8.

[4] J. Fuller, L. Stevens, and S.-L. Wang, "Risk factors for cardiovascular mortality and morbidity: The WHO multinational study of vascular disease in diabetes.," *Diabetologia*, vol. 44, no. S2, pp. S54–S64, Sep. 2001, doi: 10.1007/pl00002940.

[5] G. S. Bhavekar and A. D. Goswami, "A hybrid model for heart disease prediction using recurrent neural network and long short term memory," *Int. J. Inf. Technol.*, vol. 14, no. 4, pp. 1781–1789, Feb. 2022, doi: 10.1007/s41870-022-00896-y.

[6] D. Peddireddy, X. Fu, H. Wang, B. G. Joung, V. Aggarwal, J. W. Sutherland, and M. B.-G. Jun, "Deep learning based approach for identifying conventional machining processes from CAD data," *Proc. Manuf.*, vol. 48, pp. 915–925, Jan. 2020, doi: 10.1016/j.promfg.2020.05.130.

[7] G. Saranya and A. Pravin, "Hybrid global sensitivity analysis based optimal attribute selection using classification techniques by machine learning algorithm," *Wireless Pers. Commun.*, vol. 127, no. 3, pp. 2305–2324, Aug. 2021, doi: 10.1007/s11277-021-08796-3.

[8] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.

[9] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[10] X. Yuan, J. Chen, K. Zhang, Y. Wu, and T. Yang, "A stable AI-based binary and multiple class heart disease prediction model for IoMT," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 2032–2040, Mar. 2022, doi: 10.1109/TII.2021.3098306.

[11] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, Apr. 2017, doi: 10.1016/j.cmpb.2017.01.004.

[12] M. S. Nawaz, B. Shoaib, and M. A. Ashraf, "Intelligent cardiovascular disease prediction empowered with gradient descent optimization," *Heliyon*, vol. 7, no. 5, May 2021, Art. no. e06948, doi: 10.1016/j.heliyon.2021.e06948.

[13] D. Zhang, Y. Chen, Y. Chen, S. Ye, W. Cai, J. Jiang, Y. Xu, G. Zheng, and M. Chen, "Heart disease prediction based on the embedded feature selection method and deep neural network," *J. Healthcare Eng.*, vol. 2021, pp. 1–9, Sep. 2021, doi: 10.1155/2021/6260022.

[14] X.-Y. Gao, A. Amin Ali, H. Shaban Hassan, and E. M. Anwar, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method," *Complexity*, vol. 2021, pp. 1–10, Feb. 2021, doi: 10.1155/2021/6663455.

[15] S. Mishra, N. K. Tiwari, K. Kumari, and V. Kumawat, "Prediction of heart disease using machine learning," in *Proc. 2nd Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, May 2023, pp. 1617–1621, doi: 10.1109/ICAAIC56838.2023.10140478.

[16] K. M. Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis," *BMC Bioinf.*, vol. 21, no. 1, pp. 1–18, Jul. 2020, doi: 10.1186/s12859-020-03626-y.

[17] K. R. Sumana, Dr. H. D. Phaneendra, and K. S. Kalpashree, "Machine learning and deep learning techniques to prognosticate heart attack in due of abnormal sleep style," *Int. Res. J. Eng. Technol.*, vol. 8, no. 8, pp. 963–965, Aug. 2021.

[18] O. E. Taylor, P. S. Ezekiel, and F. B. D. Okuchaba, "A model to detect heart disease using machine learning algorithm," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 11, pp. 1–5, Nov. 2019, doi: 10.26438/ijcse/v7i11.15.

[19] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[20] A. K. Gárate-Escamila, A. H. E. Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Informat. Med. Unlocked*, vol. 19, Jan. 2020, Art. no. 100330, doi: 10.1016/j.imu.2020.100330.

[21] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.

[22] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informat. Med. Unlocked*, vol. 16, Jan. 2019, Art. no. 100203, doi: 10.1016/j.imu.2019.100203.

[23] I. D. Mienye, Y. Sun, and Z. Wang, "Improved sparse autoencoder based artificial neural network approach for prediction of heart disease," *Informat. Med. Unlocked*, vol. 18, Jan. 2020, Art. no. 100307, doi: 10.1016/j.imu.2020.100307.

[24] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *BioMed Res. Int.*, vol. 2020, pp. 1–10, Apr. 2020, doi: 10.1155/2020/9816142.

[25] J. Maiga, G. G. Hungilo, and Pranowo, "Comparison of machine learning models in prediction of cardiovascular disease using health record data," in *Proc. Int. Conf. Informat., Multimedia, Cyber Inf. Syst. (ICIMCIS)*, Jakarta, Indonesia, Oct. 2019, pp. 45–48, doi: 10.1109/ICIMCIS48181.2019.8985205.

[26] N. Hasan and Y. Bao, "Comparing different feature selection algorithms for cardiovascular disease prediction," *Health Technol.*, vol. 11, no. 1, pp. 49–62, Oct. 2020, doi: 10.1007/s12553-020-00499-2.

[27] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Informat. Med. Unlocked*, vol. 26, Jan. 2021, Art. no. 100655, doi: 10.1016/j.imu.2021.100655.

[28] V. R. Cheekati, Dr. D. Natarajasivan, and Dr. S. Indraneel, "Ensemble approaches can aid in the early detection of coronary heart disease," *NVEO-Natural Volatiles Essential Oils J.*, vol. 8, no. 5, pp. 12224–12239, Dec. 2021.

[29] B. Martins, D. Ferreira, C. Neto, A. Abelha, and J. Machado, "Data mining for cardiovascular disease prediction," *J. Med. Syst.*, vol. 45, no. 1, pp. 1–8, Jan. 2021, doi: 10.1007/s10916-020-01682-8.

[30] S. Ouf and A. I. B. ElSeddawy, "A proposed paradigm for intelligent heart disease prediction system using data mining techniques," *J. Southwest Jiaotong Univ.*, vol. 56, no. 4, pp. 220–240, Aug. 2021, doi: 10.35741/issn.0258-2724.56.4.19.

[31] R. Hagan, C. J. Gillan, and F. Mallett, "Comparison of machine learning methods for the classification of cardiovascular disease," *Informat. Med. Unlocked*, vol. 24, Jan. 2021, Art. no. 100606, doi: 10.1016/j.imu.2021.100606.

[32] M. J. Ali, B. C. Das, S. K. Saha, A. A. Biswas, and P. Chakraborty, "A comparative study of machine learning algorithms to detect cardiovascular disease with feature selection method," in *Machine Intelligence and Data Science Applications* (Lecture Notes on Data Engineering and Communications Technologies). Singapore: Springer, Aug. 2022, pp. 573–586, doi: 10.1007/978-981-19-2347-0_45.

[33] C. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, Feb. 2023, doi: 10.3390/a16020088.

[34] K. Mahalakshmi and P. Sujatha, "Critical analysis of feature selection methods for data preprocessing with heart disease dataset," in *Data Intelligence and Cognitive Informatics*. Singapore: Springer, 2022, pp. 667–682, doi: 10.1007/978-981-16-6460-1_51.

[35] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100060, doi: 10.1016/j.health.2022.100060.

[36] M. Kavitha, S. Srinivasulu, M. M. Reddy, V. Gopikrishna, S. P. Kumar, and S. Kavitha, "Hybrid model using feature selection and classifier in big data healthcare analytics," in *Inventive Communication and Computational Technologies* (Lecture Notes in Networks and Systems). Singapore: Springer, 2022, pp. 777–791, doi: 10.1007/978-981-16-5529-6_59.

[37] H. A. Mengash, L. Hussain, H. Mahgoub, A. Al-Qarafi, M. K. Nour, R. Marzouk, S. A. Qureshi, and A. M. Hilal, "Smart cities-based improving atmospheric particulate matters prediction using chi-square feature selection methods by employing machine learning techniques," *Appl. Artif. Intell.*, vol. 36, no. 1, May 2022, Art. no. 2067647, doi: 10.1080/08839514.2022.2067647.

[38] R. R. Sanni and H. S. Guruprasad, "Analysis of performance metrics of heart failured patients using Python and machine learning algorithms," *Global Transitions Proc.*, vol. 2, no. 2, pp. 233–237, Nov. 2021, doi: 10.1016/j.gltp.2021.08.028.

[39] A. Sommer and G. Twig, "The impact of childhood and adolescent obesity on cardiovascular risk in adulthood: A systematic review," *Current Diabetes Rep.*, vol. 18, no. 10, pp. 1–6, Aug. 2018, doi: 10.1007/s11892-018-1062-9.

[40] D. DeMers and D. Wachs, *Physiology, Mean Arterial Pressure*. Treasure Island, FL, USA: StatPearls, Jan. 2021.

[41] M. Pareek, M. Vaduganathan, T. Biering-Sørensen, C. Byrne, A. Qamar, Z. Almarzooq, A. Pandey, M. H. Olsen, and D. L. Bhatt, "Pulse pressure, cardiovascular events, and intensive blood-pressure lowering in the systolic blood pressure intervention trial (SPRINT)," *Amer. J. Med.*, vol. 132, no. 6, pp. 733–739, Jun. 2019, doi: 10.1016/j.amjmed.2019.01.001.

**D. YASO OMKARI** received the bachelor's and master's degrees in engineering from Jawaharlal Nehru Technological University Anantapur. She is currently pursuing the Ph.D. degree in computer science with VIT-AP University, Andhra Pradesh. Her current research interests include data mining, big data, nature-inspired algorithms and machine learning, and optimization techniques.

**KAREEMULLA SHAIK** received the B.Tech. degree from JNTU Hyderabad, the M.Tech. degree from JNTU Kakinada, and the Ph.D. degree from KLEF. He is currently a Senior Assistant Professor with the School of Computer Science and Engineering, VIT-AP University, Andhra Pradesh. He has over 15 years of experience in teaching. He has published and presented his work in various reputed journals and conferences and guided many UG and PG projects. He also published Indian patents. His research interests include computer networks and security, machine learning, and deep learning. He also presented many talks and organized conferences.

● ● ●