*I.J. Image, Graphics and Signal Processing,* **2024, 3, 1-14**
Published Online on June 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2024.03.01

# Machine Learning Based Decision Support System for Coronary Artery Disease Diagnosis

**Şükrü Alkan***
Master's Degree, Sakarya University, Faculty of Engineering, 54187, Serdivan, Sakarya, Turkey
Email: sukru.alkan@ogr.sakarya.edu.tr
ORCID iD: https://orcid.org/0000-0003-1904-3150
*Corresponding Author

**Muhammed Kürşad UÇAR**
Associate professor, Sakarya University, Faculty of Engineering, 54187, Serdivan, Sakarya, Turkey
Email: mucar@sakarya.edu.tr
ORCID iD: https://orcid.org/0000-0002-0636-8645

Received: 14 March, 2023; Revised: 23 April, 2023; Accepted: 12 February, 2024; Published: 08 June, 2024

**Abstract:** Coronary artery disease (CAD) causes millions of deaths worldwide every year. The earliest possible diagnosis is quite important, as in any diseases, for heart diseases causing such a large amount of death. The diagnosis processes have been more successful thanks to the recent studies in medicine and the rapid improvement in computer sciences. In this study, the goal is to employ machine learning methods to facilitate rapid disease diagnosis without the need to observe negative outcomes. The dataset utilized in this study was obtained from an IEEE DataPort data repository. The dataset consists of two classes. Firstly, new features have been produced by using the features in the dataset. Then, datasets that consist of multiple features have been created by using feature selection algorithms. Three models, specifically Support Vector Machines (SVM), the k-Nearest Neighbor algorithm (kNN), and Decision Tree ensembles (EDT), were trained using custom datasets. A hybrid model has been created and the performances have been compared with the other models by using these models. The best performance has been obtained from SVM and its seven performance criteria in order of accuracy, sensitivity, specificity, F- measurement, Kappa and AUC are 97.82, 0.97, 0.99, 0.98, 0.96 and 0.98%. In summary, when evaluating the performance of the constructed models, it has been demonstrated that these recommended models could aid in the swift prediction of coronary artery disease in everyday life.

**Index Terms:** Coronary Artery Disease, Hybrid Artificial Intelligence, Machine Learning

## 1. Introduction

Coronary artery disease (CAD) emerges as a result of stenosis or blockage in veins feeding the heart muscle. Although there are important efforts about diagnosis and prevention of the disease, it has a high death rate worldwide [1]. While it is acknowledged that CAD is associated with various genetic factors, a comprehensive understanding of behavioral risk factors remains elusive. In addition, the likelihood of prognosis and disability are high in patients. It is established that behavioral risk factors such as alcohol consumption, smoking, poor nutrition, obesity, and physical inactivity predominantly elevate the risk of complications [2]. Epidemic diseases increase the death risk. This disease develops and manifests over an extended period of time [3]. The likelihood of successful treatment increases when the disease is detected in its early stages.

Electrocardiogram (ECG) and Effort tests are commonly used methods in detecting the disease. Under normal circumstances, an ECG test does not offer information about the disease. However, it aids in the detection of heart attacks or chest pain when they begin [4]. Therefore, doctors usually prefer Effort tests. However, the accuracy is not 100% in this test. In addition, CAD can be diagnosed with Computed tomography angiography, known as virtual angiography among people [4,5]. The accuracy rate is quite high but CT Angiography is mostly applied to people in risk groups. Therefore, the detection of disease can be delayed in people who are outside the risk group. Considering these issues, it becomes evident that there is a requirement for a swift and readily accessible method. Machine learning (ML) techniques are commonly used which are not rapid, cost-effective and invasive methods in detecting CAD in literature [6]. However, published results in ML-based CAD diagnosis differ significantly in terms of datasets analyzed,

characteristics, performance measures, and ML techniques applied. In this study, a new approach is recommended by doing extensive evaluation for CAD detection.

## 2. Literature Review

This section will encompass a review of various aspects of studies in the literature concerning CAD detection. Visualization based methods have been developed for detecting CAD in literature. Supervised learning-based classification was conducted using convolutional neural networks (CNN) models with single-photon emission computed tomography (SPECT) images [7]. The algorithm's readability is quite challenging. However, a classification method has been developed with the help of decision trees and convolutional neural networks by using cardiac magnetic resonance (CMR) [8]. Training a CNN requires substantial computational resources and time. Moreover, when working with a limited dataset, it can lead to the issue of overfitting. In addition to these studies, Attention-based nested U-Net and VGG-16 based classification were made by segmenting X-Ray angiography images [9]. These studies involve lengthy training processes despite their high accuracy.

Different from methods based on visualization data, methods based on medical data have also been developed. The classification was made by machine learning based particle swarm optimization by using clinical data features selection [10]. The Fisher's feature selection method algorithm was utilized to identify more discriminative feature sets, with the goal of enhancing the performance of the proposed model. In addition, a decision tree has been developed by using the C4.5 data mining algorithm which is developed for CAD detection [11]. In addition to these methods, it was aimed to detect CAD by using fuzzy decision support systems based on rough set theory [12]. Instead of creating a single model on a dataset, five different models - decision tree, random forest, support vector machine, adaptive boosting and gradient boosting - were trained on the same data set and their performances were compared and the best classification method was tried to be determined [13]. In this study, a combination of Recursive Feature Elimination and Boruta feature selection methods was utilized to enhance the discriminative capability of CAD diagnosis. For the diagnosis of CAD with the same approach; logistic regression and discriminant analysis methods were used [14] and in addition to this, the best classification method was determined by using SVM, naive bayes and kNN methods [15]. The accuracy of these studies is lower compared to similar studies in the literature. Moreover, it has been tried to diagnose CAD with random forest, logistic regression and support vector machine methods [16]. With the same approach ten different models' performances were compared on the same dataset [17]. The disease was classified with different artificial intelligence models by selecting the features according to the Gini index and gain ratio weight of the features in the dataset [18]. Cost-effective and non-invasive techniques have been employed for CAD diagnosis, encompassing electrocardiogram-based analysis, health data evaluation, and the analysis of heart sounds.

Dataset based hybrid models were created by using more than one dataset rather than a single dataset, by training the same models from each dataset [19]. Again, with the same approach nine different models were trained on two different datasets and their performances were compared [20]. Apart from using only the medical data or visualization-based data, CAD classification was done with different deep learning methods by using both SPECT images and medical data with a different approach [21]. The distinctive aspect of this study, in comparison to existing literature, lies in the breadth of the dataset; however, the primary challenge in achieving rapid CAD prediction remains the collection of this data.

As a result of literature research, while the rate of accuracy obtained from studies ranged from 80-100%, visualization data or clinical information were generally classified by using artificial intelligence algorithms. The most effective among current solutions are studies based on datasets containing image-related data. The systems making classification by using visualization based clinical information were complex methods whose machine learning process consumed a long time. However, a system developed by using only clinical information will be much faster and easily accessible. The results differ in the systems developed by using clinical information. When all these conditions were considered, it was observed that a new method was needed.

The main problem to be tackled in this study is the absence of early diagnosis of coronary artery disease in individuals who do not fall into the high-risk category. The intended solution here is to enable the diagnosis of the disease in individuals outside the risk group solely based on clinical findings when EKG or Effort Tests fail to provide a diagnosis. Hence, datasets previously gathered from individuals diagnosed with CAD, along with artificial intelligence methods, were employed. Feature extraction methods were utilized to enhance the dataset's features and improve the performance of the generated models. Additionally, all features were ranked according to the Eta Correlation Coefficient, resulting in the creation of datasets with varying feature compositions. Subsequently, four different models were trained with these generated datasets, and their performances were compared. The use of the Feature Extraction method and the feature selection algorithm based on the Eta Correlation Coefficient have proven to enhance the performance of the models. The average performance of the models developed at the conclusion of the study indicated an accuracy of 88.59%, a sensitivity of 0.86, a specificity of 0.91, and an F-value of 0.88. The results show that the system can deliver coronary artery disease better than the conventional methods.

## 3. Materials and Methods

Application steps in this study are summarized in Fig. 1. In the first step, 25 characteristic features were extracted for each sample by using 11 features in the dataset. Characteristic features extracted particularly in the feature selection stage were created according to the features and correlation between labels and coefficient of openness in the original dataset. As for the last stage, the performances of feature groups which were determined in each stage of feature extraction, were assessed by using a determined classification algorithm. In addition, thereby using the determined classification algorithm, hybrid models were created. Finally, coronary artery disease was detected via created models. The ease of data collection and the algorithm's applicability have facilitated the achievement of the research goals.

### *3.1. Data Collection*

The data set used in our study was taken from Liverpool Moore's University based Institute of Electrical and Electronics Engineers Dataport (IEEE Dataport) [22]. This dataset consists of Cleveland, Hungarian, Switzerland, Statlog (Heart) Data Set and Long Beach VA which has 5 common heart diseases. 1190 samples having 11 common features from these 5 heart diseases data sets, take part as a result of this combination [23]. 23.6% of patients are women and 76.4% of them are men in the data set. The used datasets were consolidated into a single dataset in order to help to improve research on machine learning algorithms about coronary artery disease, to improve clinical diagnosis and early diagnosis. Dataset distribution is shown in Table 1.

Table 1. Distribution of the dataset

| | **DN** | **DP** |
|---|---|---|
| **Normal** | 561 | 47.14% |
| **Heart Disease** | 629 | 52.86% |
| **Total** | 1190 | 100% |

DN: Number of Data,
DP: Percentage of Data

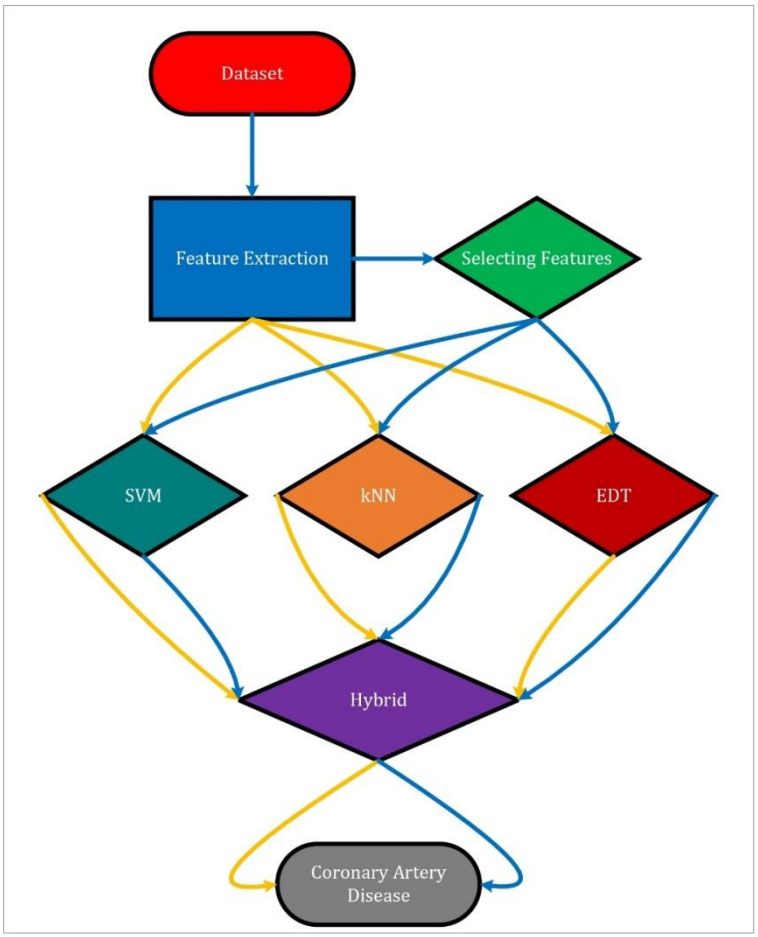



Fig. 1. Application flowchart

### *3.2. Feature Extraction*

25 features were extracted from 11 clinical data in the dataset. The formulas belonging to feature extraction are shown in Table 1 and the features denoted with '*' were computed using the MATLAB library [24]. The number of features were 36 in total. Mathematical and code representation of the features are shown in Table 2.

Table 2. Representation of features mathematical and code

| No | Features | Equation |
|---|---|---|
| **1** | Kurtosis | $x_{kur} = \frac{\sum_{i=1}^{n}\left(x(i)-\bar{x}\right)^4}{(n-1)S^4}$ |
| **2** | Skewness | $x_{ske} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)}{(n-1)S^3}$ |
| **3** | * IQR | $IQR = iqr(x)$ |
| **4** | CV | $CV = \left(S/\bar{x}\right)100$ |
| **5** | Geometric Mean | $G = \sqrt[n]{x_1+...+x_n}$ |
| **6** | Harmonic Mean | $H = n/\left(\frac{1}{x_1}+...+\frac{1}{x_n}\right)$ |
| **7** | Activity - Hjort Parameters | $A = S^2$ |
| **8** | Mobility - Hjort Parameters | $M = S_1^2/S^2$ |
| **9** | Complexity - Hjort Parameters | $C = \sqrt{\left(S_2^2/S_1^2\right)^2-\left(S_1^2-S^2\right)^2}$ |
| **10** | * Maximum | $x_{\max} = \max(x_i)$ |
| **11** | Median | $x = \begin{cases} x_{\frac{n+1}{2}} & : x\,odd \\ \frac{1}{2}\left(x_{\frac{n}{2}}+x_{\frac{n}{2}+1}\right) & : x\,even \end{cases}$ |
| **12** | * Mean Absolute Deviation | $MAD = mad(x)$ |
| **13** | * Minimum | $x_{\min} = \min(x_i)$ |
| **14** | * Central Moments | $CM = moment(x,10)$ |
| **15** | Mean | $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} = \frac{1}{n}\left(x_1+...+x_n\right)$ |
| **16** | Average Curve Length | $CL = \frac{1}{n}\sum_{i=2}^{n}\lvert x_i-x_{i-1}\rvert$ |
| **17** | Average Energy | $E = \frac{1}{n}\sum_{i=1}^{n}x_i^2$ |
| **18** | Root Mean Squared | $X_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}$ |
| **19** | Standard Error | $S_{\bar{x}} = S/\sqrt{n}$ |
| **20** | Standard Deviation | $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)}$ |
| **21** | Shape Factor | $SF = X_{rms}/\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i\rvert}\right)$ |
| **22** | * Singular Value Decomposition | $SVD = svd(x)$ |
| **23** | * 25% Trimmed Mean | $T25 = trimmean(x,25)$ |
| **24** | * 50% Trimmed Mean | $T50 = trimmean(x,50)$ |
| **25** | Average Teager Energy | $TE = \frac{1}{n}\sum_{i=3}^{n}\left(x_{i-1}^2-x_i x_{i-2}\right)$ |

* The property was computed using MATLAB
IQR Interquartile Range, CV Coefficient of Variation
$S^2$ : variance of the signal $x$
$S_1^2$ : Variance of the 1st derivative of the signal $x$
$S_2^2$ : Variance of the 2nd derivative of the signal $x$

### *3.3. Features Selection/Sort Algorithm*

Eta correlation coefficient-based feature selection algorithm was used in this study.

#### *3.3.1. Eta Correlation Coefficient Based Feature Selection Algorithm*

In the realm of machine learning, diverse data types exist, including class labels which frequently represent unordered qualitative variables. Eta correlation coefficient ( $r_{pb}$ ) is the openness and correlation coefficient between the features and the labels (1) [25].

$$r_{pb} = \frac{\bar{Y}_1 - \bar{Y}_0}{s_y} \sqrt{p_0 p_1} \tag{1}$$

Within the equation, $\bar{Y}_0$ and $\bar{Y}_1$ denote the means of data within class 0 and class 1, respectively. Meanwhile, $s_y$ represents the standard deviation calculated across all data points encompassing both classes (2).

$$s_y = \sqrt{\frac{\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}}{n}} \tag{2}$$

$N$ , $N_0$ , and $N_1$ correspond to the total number of elements, the count within Class 0, and the count within Class 1, respectively. Equation (3) illustrates the values for $p_0$ and $p_1$ .

$$p_0 = \frac{N_0}{N}, p_1 = \frac{N_1}{N} \tag{3}$$

It is used while calculating correlation coefficient between qualitative and continuous variables. When data type changes, the correlation calculation type also changes. It indicates the direction of the relationship which exists among the extracted features [25]. This value is between 0-1 and if the result is found close to 1 value, it is said that there is a positive direction between the features.

When the Eta correlation coefficient is calculated, Eta values are firstly classified from largest to smallest. The features and Eta values are shown in Table 3. The percentage of features which are going to be selected is determined. By starting with 5% of features, 20 datasets are formed by increasing the features by fives. The dataset and feature numbers are displayed in Table 4. The objective here is to evaluate the system with varying numbers of features, and a performance assessment was conducted based on these tests.

Table 3. Correlation values of features

| No | FI | ES | No | FI | ES |
|---|---|---|---|---|---|
| **1** | 16 | 0.5056 | **19** | 18 | 0.1850 |
| **2** | 36 | 0.5056 | **20** | 22 | 0.1850 |
| **3** | 34 | 0.4815 | **21** | 2 | 0.1824 |
| **4** | 28 | 0.4601 | **22** | 19 | 0.1743 |
| **5** | 11 | 0.4159 | **23** | 20 | 0.1743 |
| **6** | 33 | 0.4133 | **24** | 5 | 0.1639 |
| **7** | 35 | 0.3984 | **25** | 6 | 0.1631 |
| **8** | 8 | 0.3436 | **26** | 1 | 0.1624 |
| **9** | 21 | 0.3346 | **27** | 25 | 0.1569 |
| **10** | 27 | 0.3113 | **28** | 4 | 0.1464 |
| **11** | 23 | 0.2844 | **29** | 17 | 0.1454 |
| **12** | 3 | 0.2696 | **30** | 7 | 0.1329 |
| **13** | 9 | 0.2677 | **31** | 29 | 0.1214 |
| **14** | 26 | 0.2620 | **32** | 10 | 0.0950 |
| **15** | 12 | 0.2322 | **33** | 13 | 0.0907 |
| **16** | 31 | 0.2167 | **34** | 24 | 0.0736 |
| **17** | 15 | 0.2118 | **35** | 32 | 0.0731 |
| **18** | 30 | 0.1984 | **36** | 14 | 0.0099 |

ES: Eta Score, FI: Feature Id

Table 4. Selected properties table

| L | FN | FP |
|---|---|---|
| **1** | 2 | 5% |
| **2** | 4 | 10% |
| **3** | 5 | 15% |
| **4** | 7 | 20% |
| **5** | 9 | 25% |
| **6** | 11 | 30% |
| **7** | 13 | 35% |
| **8** | 14 | 40% |
| **9** | 16 | 45% |
| **10** | 18 | 50% |
| **11** | 20 | 55% |
| **12** | 22 | 60% |
| **13** | 23 | 65% |
| **14** | 25 | 70% |
| **15** | 27 | 75% |
| **16** | 29 | 80% |
| **17** | 31 | 85% |
| **18** | 32 | 90% |
| **19** | 34 | 95% |
| **20** | 36 | 100% |

L: Level, FN: Number of Feature, FP: Percentage of Feature

### *3.4. Machine Learning*

Machine learning is a field within computer and artificial intelligence science that progressively enhances its accuracy by concentrating on constructing learning systems based on the data they are provided. It is separated into 3 groups as supervised, unsupervised and semi supervised according to the approach that is used. Supervised learning is a machine learning algorithm that involves training the model using labeled data [26]. In this model, the accuracy can be controlled because labelled input and output data are used. Examples of classification and regression algorithms can be provided. Unsupervised learning is based on unlabeled input and output data training [26]. In this model, the model has to operate on its own for realization of learning. Because the data are not labelled, the results are not completely accurate or reliable. Grouping algorithms can be given as examples. Nevertheless, semi-supervised learning is trained using a limited amount of labeled data alongside a substantial volume of unlabeled data within our dataset [27]. In this model, training takes place by employing a combination of both supervised and unsupervised learning algorithms. Genetic sorting can be provided as an example.

The machine learning algorithms employed in this study include Support Vector Machines (SVM), k-Nearest Neighbor Algorithms (kNN), and Decision Tree Ensembles (DTE). A hybrid model was created by using these three models. It was preferred because the training duration was short and the rate of accuracy was high [28]. These methods are frequently used algorithms because of their success in the literature [29–31]. The models' performances were assessed according to the various criteria. The model which gave the best result was determined according to the assessment.

50% of data were used in training and 50% of it were used during the testing process. The quantity of data utilized in both the training and testing processes is presented in Table 5.

Table 5. Training and testing distribution

| Dataset | Train (50%) | Test (50%) | Total |
|---|---|---|---|
| **Normal** | 278 | 283 | 561 |
| **Heart Disease** | 317 | 312 | 629 |

### *3.4.1. Support Vector Machines*

Support Vector Machines (SVM) are used in solving classification and regression problems. It has a double layer feeding neural net and it is a learning method based on statistics learning theory. The rationale behind the high performance of Support Vector Machines lies in the superiority of the structural risk minimization principle upon which it is founded, in contrast to the empirical risk minimization principle employed by traditional neural networks [32,33]. It is the main aim of Support Vector Machines to have the most appropriate hyperplane which tells apart two data by maximizing the distance among the support vectors in different classes [32].

### *3.4.2. K- The Nearest Neighbor Algorithm*

K- The Nearest Neighbor (kNN) Algorithm is used in solving classification problems. The algorithm performs classification by leveraging the extracted features during classification and by assessing the proximity of the data to be classified with respect to the k value of previous data points [29,34]. Euclidean distance, cosine similarity measure, Minkowski, correlation and chi-square methods are used while doing proximity calculation [34]. According to the k value determined while classifying a data, the closest k of the old classified data is taken. The main purpose of the K-The Nearest Neighbor Algorithm is to make classification by finding the nearest spots to the newest spots with closeness calculation.

### *3.4.3. Decision Tree Ensembles*

Decision Tree is a method which is used to help to determine the branch having the highest possibility to reach a leaf [33]. Data set in model learning is separated into subsets according to the various features. This process recursively continues until the classes are determined. It will rarely lead to a good generalization to use a single decision tree. More than one decision tree can be combined in order to make a better prediction. It is frequently used in solving classification and regression problems [35,36]. This method is known as "Ensembles". There are two different methods as Bagging and Boosting in order to create Decision Tree Ensembles (DTE) [35,36]. These methods have different advantages and disadvantages. The main aim of Decision Tree Ensembles is to increase the classification accuracy by using different decision trees.

### *3.4.4. Hybrid Machine Learning*

Several algorithms used in machine learning are good in solving problems such as regression and classification. Nevertheless, it is good, artificial intelligence does not emerge to its full potential. However, hybrid machine learning is machine learning which is used by combining existing methodologies [37–39]. A single machine learning algorithm is not appropriate to solve several problems. Hybrid machine learning is constitutively used with the aim of improving a new model and covering up other's deficiency [37–39]. In this context, a hybrid model was obtained by using the average method of three different models used in the study. The mathematical expression of the hybrid model is expressed in equation (4).

$$\overline{x}_{hybrid} = \frac{\sum x}{n} \tag{4}$$

$\overline{x}_{hybrid}$ represents the mean of the predicted class values derived from a set of machine learning models utilized in this study. These predicted class values are denoted as $x$ values. The value of $n$ corresponds to the number of $x$ values, which in this case is 3. $\sum x$ , on the other hand, stands for the sum of predicted class values from all these machine learning models, which can be calculated as $\sum x = x_{SVM} + x_{kNN} + x_{DTE}$ .

### *3.5. Performance Evaluation Criteria*

6 different performance evaluation criteria - accuracy, sensitivity, specificity, F measurement, Kappa and area (AUC) under the receiver operating characteristic (ROC) - were used within the scope of study.

### *3.5.1. Accuracy*

Accuracy is the ratio of the data that the model predicts correctly to the total dataset. The value is between the range of 0-100. The closer the predicted values align with the actual values, the higher the accuracy value becomes.

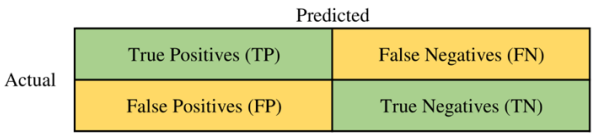



Fig. 2. Confusion matrix

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100 \tag{5}$$

*3.5.2. Sensitivity*

Sensitivity indicates how much of the data which should be positively predicted by the model, is correctly detected. In other words, it states how much of the actual positive data is correctly predicted.

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

*3.5.3. Specificity*

Specificity shows how much of data which should be negatively predicted by the model, is correctly detected. In other words, it states how much of the actual negative data is correctly predicted.

$$Specificity = \frac{TN}{FP + TN} \tag{7}$$

*3.5.4. F Score*

The F-Score is calculated as the harmonic mean of the sensitivity and specificity values of the model. It specifies model effectiveness. The value is between 0-1. The performance increases when the value approaches 1.

$$F = 2 * \frac{Sensitivity * Specificity}{Sensitivity + Specificity} \tag{8}$$

*3.5.5. Kappa*

Kappa is the aleatoric coefficient of concordance. It gives information about the reliability of the model. Different value ranges take part in the literature. The value range used in this study is in the range of [-1,1]. It is calculated with equation (9). The $p_0$ in the formula is the accuracy value of the model. However, $p_e$ is the measurement of harmony between the model's prediction and actual values and it is calculated with equation (10).

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

$$p_e = p_{e1,target} * p_{e1,predict} + p_{e2,target} * p_{e2,predict} \tag{10}$$

*3.5.6. Area Under the Curve (AUC)*

Area Under the Curve (AUC) shows how correct the model distinguishes the classes. This is expressed through the Area Under the Receiver Operating Characteristic (ROC) curve, which is a probability curve plotted against false positive values in relation to true positive values. The discrimination performance of the model is directly proportional to the AUC value.

## 4. Results

The aim of the study is to make detection of coronary artery disease with high accuracy, fast and easily accessible with the developed method. A data set which was created from real data was used for this method. There are 11 features and a disease class in the data set. Feature extraction is executed using the pre-existing features within the dataset. It has been noted that the model's performance improves when these extracted features are incorporated into the dataset. After that the features are classified by feature selection/extraction algorithm according to the relation level with classes (Table 3). 20 different data sets are created by selecting classified features according to relationship level (Table 4). Four different machine learning models, SVM, kNN, DTE and Hybrid, were created by using these data sets. We

divided the dataset into training and testing sets to ensure the reliability and accuracy of the results. Additionally, we evaluated the created models using seven different performance criteria. Classification performance differs according to the created model and the members of features in the data set (Table 6, Table 7). The more the number of the features increases the more the model's accuracy in the data set created by using the Eta feature extraction algorithm increases. While the hybrid model showed high performance in models created with datasets containing few features, the SVM model showed higher performance as the number of features increased. Machine learning models may exhibit varying performances even when the number of features is identical. Consequently, it was observed that the performance of the hybrid model might be lower than that of the other models employed.

The change of performance criteria for each model and feature numbers are summarized in Fig 4 in order to make tables more understandable. Moreover, the radar which represents model performance evaluation criteria created for each data set is shown in Fig 3. Each data set's performance values are good depending on how close they are to radar's outside. It was found that as a result of developed models, 20, 19, 17 numbered feature selection levels have the highest prediction performance with SMV. However, the hybrid model is seen to be more successful in 1, 4, 5, 7- 16 feature selection levels which have a couple of features. Even if the accuracy levels show changes in other levels, the developed system is practical. When evaluating the obtained results, it is desired that the accuracy level value be near 100, and the sensitivity, specificity, F-measurement, and kappa values should approach 1 in order to identify the best model. When considering models created from a dataset containing different number features, it is understood that the best model is SMV set (Table 6, Table 7).

Table 6. Heart disease prediction models

| **Info** | | | | **Performance Evaluation Criteria** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **L** | **FN** | **FP** | **Model** | **Accuracy** | **Sensitivity** | **Specificity** | **F Score** | **Kappa** | **AUC** |
| **1** | 2 | 5 | SVM | 79.66 | 0.76 | 0.83 | 0.79 | 0.59 | 0.79 |
| | | | kNN | 79.66 | 0.76 | 0.83 | 0.79 | 0.59 | 0.79 |
| | | | Ensemble | 79.66 | 0.76 | 0.83 | 0.79 | 0.59 | 0.79 |
| | | | **Hybrid** | **79.66** | **0.76** | **0.83** | **0.79** | **0.59** | **0.79** |
| **2** | 4 | 10 | SVM | 82.18 | 0.77 | 0.87 | 0.82 | 0.64 | 0.82 |
| | | | kNN | 83.03 | 0.78 | 0.88 | 0.83 | 0.66 | 0.83 |
| | | | **Ensemble** | **83.70** | **0.87** | **0.81** | **0.84** | **0.67** | **0.84** |
| | | | Hybrid | 82.86 | 0.80 | 0.86 | 0.83 | 0.66 | 0.83 |
| **3** | 5 | 15 | SVM | 81.51 | 0.75 | 0.88 | 0.81 | 0.63 | 0.81 |
| | | | kNN | 82.86 | 0.78 | 0.87 | 0.82 | 0.66 | 0.83 |
| | | | **Ensemble** | **84.20** | **0.80** | **0.88** | **0.84** | **0.68** | **0.84** |
| | | | Hybrid | 83.53 | 0.78 | 0.89 | 0.83 | 0.67 | 0.83 |
| **4** | 7 | 20 | SVM | 83.87 | 0.81 | 0.87 | 0.84 | 0.68 | 0.84 |
| | | | kNN | 86.55 | 0.86 | 0.87 | 0.87 | 0.73 | 0.87 |
| | | | Ensemble | 87.23 | 0.86 | 0.89 | 0.87 | 0.74 | 0.87 |
| | | | **Hybrid** | **87.90** | **0.86** | **0.90** | **0.88** | **0.76** | **0.88** |
| **5** | 9 | 25 | SVM | 84.54 | 0.82 | 0.87 | 0.84 | 0.69 | 0.84 |
| | | | kNN | 87.73 | 0.86 | 0.90 | 0.88 | 0.75 | 0.88 |
| | | | Ensemble | 89.08 | 0.87 | 0.91 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **89.24** | **0.86** | **0.92** | **0.89** | **0.78** | **0.89** |
| **6** | 11 | 30 | SVM | 85.21 | 0.78 | 0.92 | 0.84 | 0.70 | 0.85 |
| | | | **kNN** | **89.92** | **0.87** | **0.93** | **0.90** | **0.80** | **0.90** |
| | | | Ensemble | 85.88 | 0.84 | 0.87 | 0.86 | 0.72 | 0.86 |
| | | | Hybrid | 89.58 | 0.86 | 0.93 | 0.89 | 0.79 | 0.89 |
| **7** | 13 | 35 | SVM | 89.08 | 0.87 | 0.91 | 0.89 | 0.78 | 0.89 |
| | | | kNN | 88.91 | 0.89 | 0.89 | 0.89 | 0.78 | 0.89 |
| | | | Ensemble | 88.91 | 0.86 | 0.92 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **91.76** | **0.89** | **0.94** | **0.92** | **0.83** | **0.92** |
| **8** | 14 | 40 | SVM | 88.24 | 0.85 | 0.91 | 0.88 | 0.76 | 0.88 |
| | | | kNN | 90.08 | 0.86 | 0.94 | 0.90 | 0.80 | 0.90 |
| | | | Ensemble | 88.07 | 0.85 | 0.91 | 0.88 | 0.76 | 0.88 |
| | | | **Hybrid** | **90.08** | **0.85** | **0.95** | **0.90** | **0.80** | **0.90** |

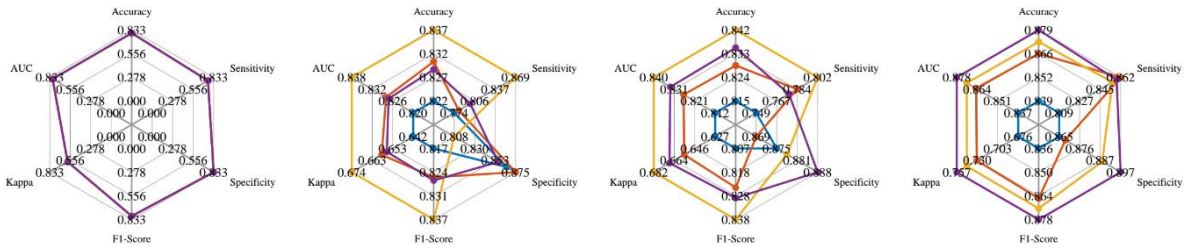
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **9** | 16 | 45 | SVM | 89.75 | 0.87 | 0.92 | 0.90 | 0.79 | 0.90 |
| | | | kNN | 90.42 | 0.86 | 0.94 | 0.90 | 0.81 | 0.90 |
| | | | Ensemble | 86.05 | 0.86 | 0.86 | 0.86 | 0.72 | 0.86 |
| | | | **Hybrid** | **90.42** | **0.88** | **0.93** | **0.90** | **0.81** | **0.90** |
| **10** | 18 | 50 | SVM | 88.57 | 0.88 | 0.89 | 0.89 | 0.77 | 0.89 |
| | | | kNN | 89.08 | 0.87 | 0.91 | 0.89 | 0.78 | 0.89 |
| | | | Ensemble | 89.24 | 0.89 | 0.90 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **91.09** | **0.89** | **0.93** | **0.91** | **0.82** | **0.91** |

L: Level, FN: Number of Feature, FP: Percentage of Feature

Table 7. Heart disease prediction models (continue)

| **Info** | | | | **Performance Evaluation Criteria** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **L** | **FN** | **FP** | **Model** | **Accuracy** | **Sensitivity** | **Specificity** | **F Score** | **Kappa** | **AUC** |
| **11** | 20 | 55 | SVM | 90.92 | 0.90 | 0.92 | 0.91 | 0.82 | 0.91 |
| | | | kNN | 90.92 | 0.89 | 0.93 | 0.91 | 0.82 | 0.91 |
| | | | Ensemble | 89.75 | 0.89 | 0.90 | 0.90 | 0.79 | 0.90 |
| | | | **Hybrid** | **92.94** | **0.92** | **0.94** | **0.93** | **0.86** | **0.93** |
| **12** | 22 | 60 | SVM | 88.74 | 0.88 | 0.89 | 0.89 | 0.77 | 0.89 |
| | | | kNN | 88.74 | 0.84 | 0.93 | 0.88 | 0.77 | 0.89 |
| | | | Ensemble | 90.25 | 0.88 | 0.93 | 0.90 | 0.80 | 0.90 |
| | | | **Hybrid** | **91.26** | **0.88** | **0.94** | **0.91** | **0.82** | **0.91** |
| **13** | 23 | 65 | SVM | 91.09 | 0.89 | 0.93 | 0.91 | 0.82 | 0.91 |
| | | | kNN | 89.58 | 0.86 | 0.93 | 0.89 | 0.79 | 0.89 |
| | | | Ensemble | 89.24 | 0.88 | 0.91 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **91.43** | **0.89** | **0.94** | **0.91** | **0.83** | **0.91** |
| **14** | 25 | 70 | SVM | 88.40 | 0.88 | 0.89 | 0.88 | 0.77 | 0.88 |
| | | | kNN | 89.08 | 0.85 | 0.93 | 0.89 | 0.78 | 0.89 |
| | | | Ensemble | 88.57 | 0.85 | 0.92 | 0.88 | 0.77 | 0.88 |
| | | | **Hybrid** | **91.26** | **0.87** | **0.95** | **0.91** | **0.82** | **0.91** |
| **15** | 27 | 75 | SVM | 88.74 | 0.88 | 0.90 | 0.89 | 0.77 | 0.89 |
| | | | kNN | 91.60 | 0.89 | 0.94 | 0.91 | 0.83 | 0.91 |
| | | | Ensemble | 90.59 | 0.89 | 0.92 | 0.91 | 0.81 | 0.91 |
| | | | **Hybrid** | **92.61** | **0.92** | **0.94** | **0.93** | **0.85** | **0.93** |
| **16** | 29 | 80 | SVM | 91.93 | 0.93 | 0.91 | 0.92 | 0.84 | 0.92 |
| | | | kNN | 89.58 | 0.87 | 0.92 | 0.89 | 0.79 | 0.89 |
| | | | Ensemble | 89.08 | 0.87 | 0.91 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **92.61** | **0.89** | **0.96** | **0.92** | **0.85** | **0.92** |
| **17** | 31 | 85 | **SVM** | **93.78** | **0.94** | **0.94** | **0.94** | **0.88** | **0.94** |
| | | | kNN | 89.41 | 0.88 | 0.91 | 0.89 | 0.79 | 0.89 |
| | | | Ensemble | 87.90 | 0.87 | 0.89 | 0.88 | 0.76 | 0.88 |
| | | | Hybrid | 91.09 | 0.89 | 0.93 | 0.91 | 0.82 | 0.91 |
| **18** | 32 | 90 | SVM | 93.61 | 0.95 | 0.93 | 0.94 | 0.87 | 0.94 |
| | | | kNN | 89.41 | 0.88 | 0.91 | 0.89 | 0.79 | 0.89 |
| | | | Ensemble | 88.91 | 0.88 | 0.90 | 0.89 | 0.78 | 0.89 |
| | | | **Hybrid** | **93.78** | **0.93** | **0.95** | **0.94** | **0.88** | **0.94** |
| **19** | 34 | 95 | **SVM** | **93.78** | **0.95** | **0.93** | **0.94** | **0.88** | **0.94** |
| | | | kNN | 88.74 | 0.85 | 0.92 | 0.88 | 0.77 | 0.89 |
| | | | Ensemble | 88.74 | 0.86 | 0.92 | 0.88 | 0.77 | 0.89 |
| | | | Hybrid | 91.93 | 0.90 | 0.94 | 0.92 | 0.84 | 0.92 |
| **20** | 36 | 100 | **SVM** | **97.82** | **0.97** | **0.99** | **0.98** | **0.96** | **0.98** |
| | | | kNN | 87.73 | 0.87 | 0.89 | 0.88 | 0.75 | 0.88 |
| | | | Ensemble | 89.75 | 0.88 | 0.91 | 0.90 | 0.79 | 0.90 |
| | | | Hybrid | 92.77 | 0.92 | 0.94 | 0.93 | 0.85 | 0.93 |

L: Level, FN: Number of Feature, FP: Percentage of Feature



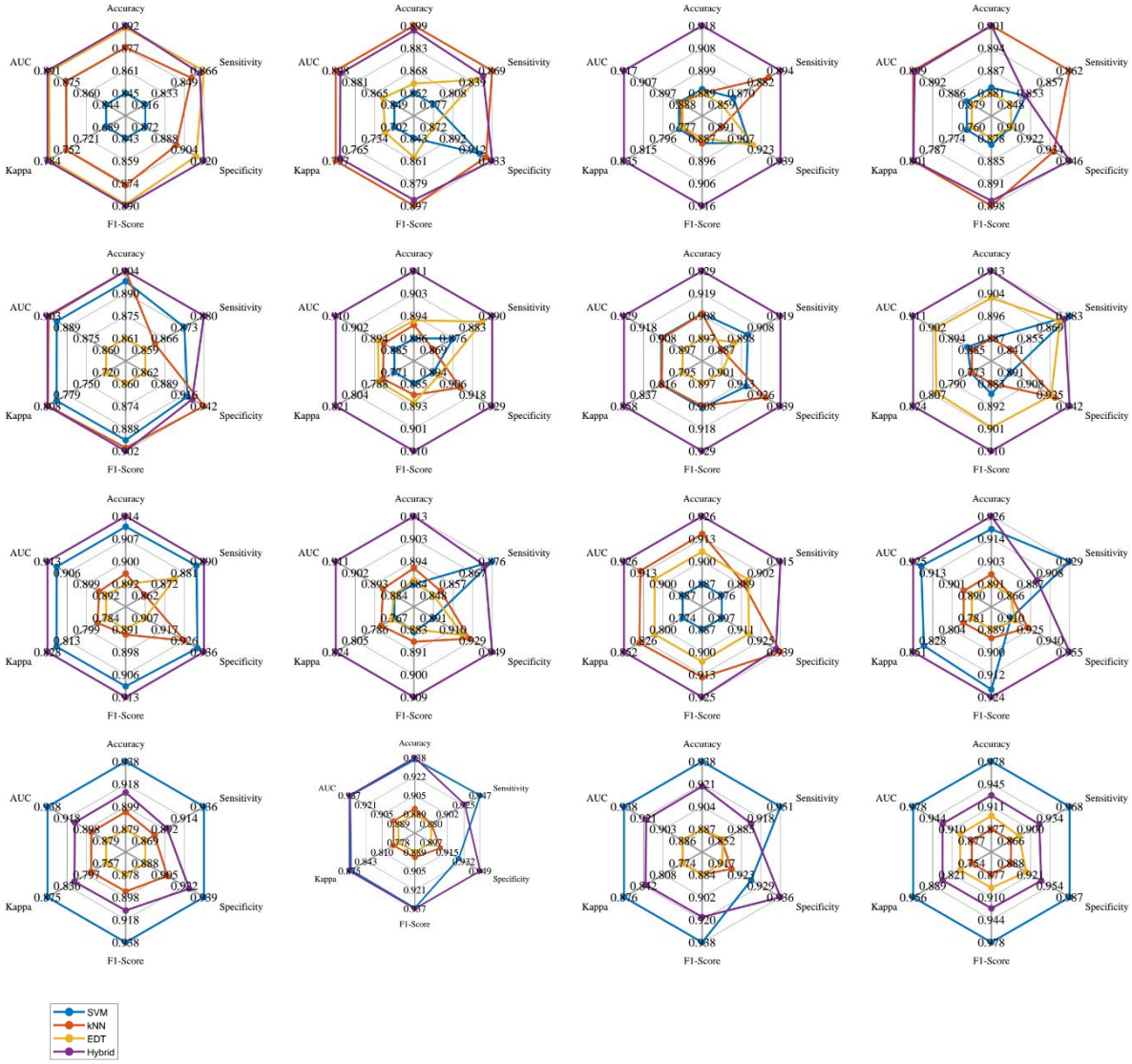



Fig. 3. Performance evaluation chart

Table 8. Comparison of current approach for coronary artery disease prediction

| Authors | Data Type | Model Used | Results |
|---|---|---|---|
| Papandrianos et al. [7] | SPECT images | CNN | accuracy of 93.33% |
| Khozeimeh et al. [8] | CMR images | CNNs | accuracy of 99.18% |
| Algarni et al. [9] | X-ray angiography | Attention-based nested U-Net and VGG-16 | accuracy of 97% |
| Shahid et al. [10] | Medical data | Particle swarm optimization based Extreme learning machine (PSO-ELM) | accuracy of 96.7% |
| Haruna et al. [11] | Medical data | Improved C4.5 data mining algorithm | accuracy of 97.23% |
| Setiawan et al. [12] | Medical data | kNN | accuracy of 92% |
| Devi et al. [13] | Medical data | Random forest | accuracy of 88% |
| Shariatnia et al. [14] | Medical data | Linear discriminant analysis | accuracy of 78.6% |
| Nassif et al. [15] | Medical data | Naive bayes | accuracy of 84% |
| Masih et al. [16] | Medical data | Deep neural network | accuracy of 96.5% |
| Tiwari et al. [17] | * Medical data | Stacked ensemble classifier | accuracy of 92.34% |
| Ghasemi et al. [18] | Medical data | Decision tree | accuracy of 99.67% |
| Doppala et al. [19] | * Medical data | Ensemble model | accuracy of 96.75% |
| Ali et al. [20] | Medical data | Decision tree | accuracy of 73.28% |
| Otaki et al. [21] | SPECT images and medical data | CNN | AUC= 83% |
| Proposed Approach | * Medical data | SVM | accuracy of 97.82% |

* The datasets used are the same

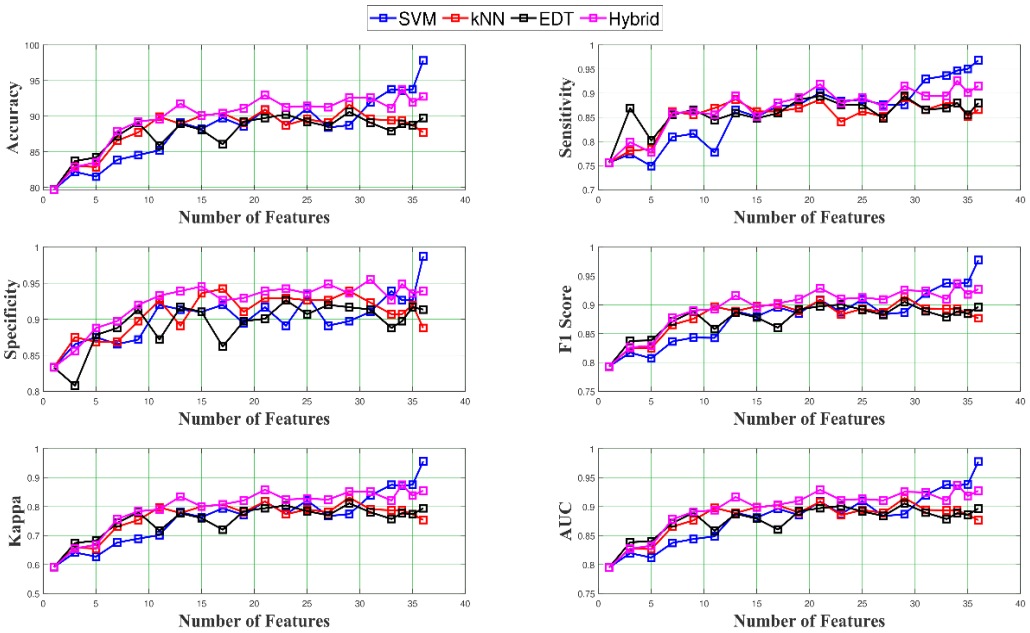



Fig. 4. Graphical summary performance indicator of heart disease prediction models

## 5. Discussion

CAD detection was explained in the previous section. The performances were presented for the data set containing a different number of features and four different models created. In this section, the study's strengths and weaknesses will be elucidated through a comparison with other studies in the literature. The CAD detection model presented in the study is based on medical data and machine learning [6]. CAD is detected from medical data taken from the individuals by benefiting from feature augmentation, feature selection and machine learning algorithm.

Early detection delays because existing methods are not highly accurate, moreover highly accurate methods can be used only after breast pain or heart attack symptoms occur or they can be applied only to patients in risk groups [4,5]. When the literature on CAD diagnosis is examined, many different methods are seen. Studies utilizing datasets composed of medical imaging techniques have predominantly favored Convolutional Neural Networks (CNN) and related algorithms [7–9,21]. Training a Convolutional Neural Network (CNN) demands significant computational resources and time. Furthermore, when dealing with a small dataset, it can result in the problem of overfitting. In some studies, a feature selection algorithm was used and it was observed that this increased the performance of the model [10,13]. In some research studies, datasets that include both health records and medical images have been utilized [21]. However, a fundamental challenge in reaching CAD predictions in these studies is the distinctiveness of the dataset in terms of its breadth, despite the difficulties associated with data collection. Comparison of existing methods about CAD is given Table 8. New methods which are highly accurate and easily accessible are tried to be developed instead of existing methods. Practical systems are seen to be developed by using existing medical data with this approach [17,19].

The most important feature of this study which is different from other diagnostic methods is to make diagnosis with machine learning methods by using direct medical data without observing any disease symptoms. With this purpose, feature extraction and eta correlation coefficient-based feature selection algorithms were used to increase the performance. Model performances were increased by creating more related features with feature extraction and our class values. Even though the success rate has partially gone down with the feature selection algorithm, processing load has been significantly decreased. The changes of the model performances have been observed in (Table 5, Table 6). In addition, Higher performance was obtained with hybrid models in low feature levels. This implementation moved ahead of literature with feature augmentation, eta correlation coefficient-based feature selection and hybrid model. When it is compared with the literature, the model is reliable and realistic.

## 6. Conclusion

The findings of this study suggest that the diagnosis of coronary artery disease (CAD) can be achieved utilizing healthcare data through machine learning and signal processing techniques. In the literature, various signals and combinations thereof are used for CAD diagnosis. However, using a non-invasive and easily accessible signal for early diagnosis is crucial for all patient groups. The medical objective of this research is to detect CAD using simple healthcare data before it becomes untreatable. The study's benefits can be elucidated as follows. The proposed model offers many advantages due to its fast diagnosis, reliability, high accuracy, technological infrastructure, and low cost.

To address the early diagnosis problem of CAD with healthcare data, three different classification algorithms were utilized, and a hybrid model was obtained by combining these models. SVM and hybrid models appear to yield remarkable results for CAD detection.

The novel contributions outlined in the findings of this study are as follows. CAD can be accurately detected using healthcare data. Feature extraction and selection processes improved model performance. Artificial intelligence-based models have enhanced system reliability. Successful models can be implemented and are believed to serve as diagnostic aids to CAD experts, offering objective and faster interpretation.

## References


[1] C. Burke, *Coronary Artery Disease: Characteristics, Management and Long-Term Outcomes* (Nova Science Publishers, New York, 2016).

[2] W. World Health Organization, *Cardiovascular Diseases (CVDs)*, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[3] P. Rola et al., *The Usefulness of the C2HEST Risk Score in Predicting Clinical Outcomes among Hospitalized Subjects with COVID-19 and Coronary Artery Disease*, Viruses 14, (2022).

[4] E. Maffei et al., *Stress-ECG vs. CT Coronary Angiography for the Diagnosis of Coronary Artery Disease: A "Real-World" Experience*, Radiol Med 115, 354 (2010).

[5] D. I. Feldman, J. Latina, J. Lovell, R. S. Blumenthal, and A. Arbab-Zadeh, *Coronary Computed Tomography Angiography in Patients with Stable Coronary Artery Disease*, Trends Cardiovasc. Med. 32, 421 (2022).

[6] R. Alizadehsani, M. Abdar, M. Roshanzamir, A. Khosravi, P. M. Kebria, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, and U. R. Acharya, *Machine Learning-Based Coronary Artery Disease Diagnosis: A Comprehensive Review*, Comput. Biol. Med. 111, 103346 (2019).

[7] N. I. Papandrianos, A. Feleki, E. I. Papageorgiou, and C. Martini, *Deep Learning-Based Automated Diagnosis for Coronary Artery Disease Using SPECT-MPI Images*, J. Clin. Med. 2022, Vol. 11, Page 3918 11, 3918 (2022).

[8] F. Khozeimeh et al., *RF-CNN-F: Random Forest with Convolutional Neural Network Features for Coronary Artery Disease Diagnosis Based on Cardiac Magnetic Resonance*, Sci. Reports 2022 121 12, 1 (2022).

[9] M. Algarni, A. Al-Rezqi, F. Saeed, A. Alsaeedi, and F. Ghabban, *Multi-Constraints Based Deep Learning Model for Automated Segmentation and Diagnosis of Coronary Artery Disease in X-Ray Angiographic Images*, PeerJ Comput. Sci. 8, e993 (2022).

[10] A. H. Shahid, M. P. Singh, B. Roy, and A. Aadarsh, *Coronary Artery Disease Diagnosis Using Feature Selection Based Hybrid Extreme Learning Machine*, Proc. - 3rd Int. Conf. Inf. Comput. Technol. ICICT 2020 341 (2020).

[11] A. A. Haruna, L. J. Muhammad, B. Z. Yahaya, E. J. Garba, N. D. Oye, and L. T. Jung, *An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease*, Proceeding 2019 Int. Conf. Digit. Landscaping Artif. Intell. ICD 2019 48 (2019).

[12] N. A. Setiawan, P. A. Venkatachalam, A. Fadzil, and M. Hani, *Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System*, Proc. Int. Conf. Man-Machine Syst. 11 (2020).

[13] K. N. Devi, S. Suruthi, and S. Shanthi, *Coronary Artery Disease Prediction Using Machine Learning Techniques*, 8th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2022 1029 (2022).

[14] S. Shariatnia, M. Ziaratban, A. Rajabi, A. Salehi, K. Abdi Zarrini, and M. Vakili, *Modeling the Diagnosis of Coronary Artery Disease by Discriminant Analysis and Logistic Regression: A Cross-Sectional Study*, BMC Med. Inform. Decis. Mak. 22, 1 (2022).

[15] A. B. Nassif, O. Mahdi, Q. Nasir, M. A. Talib, and M. Azzeh, *Machine Learning Classifications of Coronary Artery Disease*, 2018 Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. ISAI-NLP 2018 - Proc. (2018).

[16] N. Masih, H. Naz, and S. Ahuja, *Multilayer Perceptron Based Deep Neural Network for Early Detection of Coronary Heart Disease*, Health Technol. (Berl). 11, 127 (2021).

[17] A. Tiwari, A. Chugh, and A. Sharma, *Ensemble Framework for Cardiovascular Disease Prediction*, Comput. Biol. Med. 146, 105624 (2022).

[18] F. Ghasemi, B. S. Neysiani, and N. Nematbakhsh, *Feature Selection in Pre-Diagnosis Heart Coronary Artery Disease Detection: A Heuristic Approach for Feature Selection Based on Information Gain Ratio and Gini Index*, 2020 6th Int. Conf. Web Res. ICWR 2020 27 (2020).

[19] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, *A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques*, J. Healthc. Eng. 2022, (2022).

[20] Z. Ali, N. Naseer, and H. Nazeer, *Cardiovascular Disease Detection Using Multiple Machine Learning Algorithms and Their Performance Analysis*, 1 (2023).

[21] Y. Otaki et al., *Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease*, JACC Cardiovasc. Imaging 15, 1091 (2022).

[22] Manu Siddhartha, *Heart Disease Dataset (Comprehensive)*, IEEE Dataport (2020).

[23] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, M. Panahiazar, A. Koohestani, F. Khozeimeh, S. Nahavandi, and N. Sarrafzadegan, *A Database for Using Machine Learning and Data Mining Techniques for Coronary Artery Disease Diagnosis*, Sci. Data 6, 1 (2019).

[24] Wallisch P, Lusignan ME, Benayoun MD, Baker TI, Dickey AS, and Hatsopoulos NG., *MATLAB for Neuroscientists : An Introduction to Scientific Computing in MATLAB*, Second edi (Elsevier, 2014).

[25] M. K. Uçar, *Eta Correlation Coefficient Based Feature Selection Algorithm for Machine Learning: E-Score Feature Selection Algorithm*, J. Intell. Syst. Theory Appl. 2, 7 (2019).

[26] Sunila Gollapudi, *Practical Machine Learning* (Packt Publishing, Birmingham, 2016).

[27] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning* (The MIT Press, 2006).

[28] R. U. Rasool, U. Ashraf, K. Ahmed, H. Wang, W. Rafique, and Z. Anwar, *Cyberpulse: A Machine Learning Based Link Flooding Attack Mitigation System for Software Defined Networks*, IEEE Access 7, 34885 (2019).
[29] I. B. Aydilek and A. Arslan, *A Hybrid Method for Imputation of Missing Values Using Optimized Fuzzy C-Means with Support Vector Regression and a Genetic Algorithm*, Inf. Sci. (Ny). 233, 25 (2013).
[30] I. B. Aydilek and A. Arslan, *A Novel Hybrıd Approach To Estımatıng Mıssıng Values In Databases Usıng K-Nearest Neıghbors And Neural Networks*, Int. J. Innov. Comput. Inf. Control ICIC Int. C 8, 4705 (2012).
[31] M. Arican and K. Polat, *Binary Particle Swarm Optimization (BPSO) Based Channel Selection in the EEG Signals and Its Application to Speller Systems*, J. Artif. Intell. Syst. 2, 27 (2020).
[32] Y. Zhao and Q. He, *An Unbalanced Dataset Classification Approach Based on V-Support Vector Machine*, Proc. World Congr. Intell. Control Autom. 2, 10496 (2006).
[33] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, *A Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine*, Proc. - 2nd Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2016 (2017).
[34] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, *The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets*, Springerplus (2016).
[35] L. Breiman, *Random Forests*, Mach. Learn. 2001 451 45, 5 (2001).
[36] L. Breiman, *Bagging Predictors*, Mach. Learn. 1996 242 24, 123 (1996).
[37] M. K. Uçar, Z. Uçar, F. Köksal, and N. Daldal, *Estimation of Body Fat Percentage Using Hybrid Machine Learning Algorithms*, Measurement 167, 108173 (2021).
[38] M. R. Bozkurt, M. K. Uçar, F. Bozkurt, and C. Bilgin, *Development of Hybrid Artificial Intelligence Based Automatic Sleep/Awake Detection*, IET Sci. Meas. Technol. 14, 353 (2020).
[39] I. Topal and M. K. Ucar, *Hybrid Artificial Intelligence Based Automatic Determination of Travel Preferences of Chinese Tourists*, IEEE Access 7, 162530 (2019).

## Authors' Profiles

**Şükrü Alkan** was born in Ankara, Turkey. He received B.Sc. in Electrical and Electronics Engineering from Kırıkkale University, Kırıkkale, Turkey in 2015; He received his MSc. degree in Electrical and Electronics Engineering in 2021 from Sakarya University, Sakarya, Turkey. His research interest includes digital signal processing, artificial intelligence, and machine learning.

**Muhammed Kürşad Uçar** was born in Gümüşhane, Turkey. He received the Electrical and Electronics Engineering degree from Mustafa Kemal University, Turkey, and the master's degree in electrical and electronic engineering and the Ph.D. degree from Sakarya University, in 2017. He is currently Associate Professor with the Department of Electrical and Electronics Engineering, Sakarya University. His research interests include biomedical signal processing, statistical signal processing, digital signal processing, artificial intelligence, and machine learning.

**How to cite this paper:** Şükrü Alkan, Muhammed Kürşad UÇAR, "Machine Learning Based Decision Support System for Coronary Artery Disease Diagnosis", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.16, No.3, pp. 1-14, 2024. DOI:10.5815/ijigsp.2024.03.01