Frontiers in Health Informatics
2024; 13: 202
www.ijmi.ir
Open Access

Enhancing coronary artery diseases screening: A comprehensive assessment of machine learning approaches using routine clinical and laboratory data

Shahryar Naji[1], Zahra Niazkhani[2,3], Kamal Khademvatan[4], Habibollah Pirnejad[3,5]*

[1]Student Research Committee, Urmia University of Medical Sciences, Urmia, Iran
[2]Nephrology and Kidney Transplant Research Center, Clinical Research Institute, Urmia University of Medical Sciences, Urmia, Iran
[3]Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, Rotterdam, The Netherlands
[4]Department of Cardiology, Urmia University of Medical Sciences, Urmia, Iran
[5]Patient Safety Research Center, Clinical Research Institute, Urmia University of Medical Sciences, Urmia, Iran

Article Info

Article type:
Research

Article History:
Received: 2023-12-03
Accepted: 2024-03-22
Published: 2024-04-13

**** Corresponding author:***
Habibollah Pirnejad

Patient Safety Research Center, Clinical Research Institute, Urmia University of Medical Sciences, Urmia, Iran

Email: pirnejad@eshpm.eur.nl

Keywords:
Coronary Artery Disease
Coronary Angiography
Machine Learning
Artificial Intelligence

A B S T R A C T

Introduction: Coronary artery disease (CAD) stands among the leading global causes of mortality, underscoring the critical necessity for early detection to facilitate effective treatment. Although Coronary Angiography (CA) serves as the gold standard for diagnosis, its limitations for screening, including side effects and cost, necessitate alternative approaches. This study focuses on the development and comparison of machine learning techniques as substitutes for CA in CAD screening, leveraging routine clinical and laboratory data.

Material and Methods: Various machine learning classification algorithms—decision tree, k-nearest neighbor, artificial neural network, support vector machine, logistic regression, and stacked ensemble learning were employed to differentiate CAD and healthy subjects. Feature selection algorithms, namely LASSO and ReliefF, were utilized to prioritize relevant features. A range of evaluation metrics, including accuracy, precision, sensitivity, specificity, AUC, F1 score, ROC curve, and NPV, were applied. The SHAP technique was employed to elucidate and interpret the artificial neural network model.

Results: The artificial neural network, support vector machine, and stacked ensemble learning models demonstrated excellent results in a 10-fold cross-validation evaluation using features selected by LASSO and ReliefF. With the LASSO feature selection algorithm, these models achieved accuracies of 90.38%, 90.07%, and 90.39%, sensitivities of 94.43%, 93.03%, and 93.96%, and specificities of 80.27%, 82.77%, and 81.52%, respectively. Using ReliefF, the accuracies were 88.79%, 88.77%, and 90.06%, sensitivities were 92.12%, 91.66%, and 93.98%, and specificities were 80.13%, 81.38%, and 80.13%, respectively. The SHAP technique revealed that typical and atypical chest pain, hypertension, diabetes mellitus, T inversion, and age were the most influential features in the neural network model.

Conclusion: The machine learning models developed in this study exhibit high potential for non-invasive screening and diagnosis of CAD in the Z-Alizadeh Sani dataset. However, further studies are essential to validate and apply these models in real-world and clinical settings.

Cite this paper as:
Naji S, Niazkhani Z, Khademvatan K, Pirnejad H. Enhancing coronary artery diseases screening: A comprehensive assessment of machine learning approaches using routine clinical and laboratory data. Front Health Inform. 2024; 13: 202. DOI: 10.30699/fhi.v13i0.555

INTRODUCTION

Coronary artery disease (CAD) is a common chronic disease in which the coronary arteries become narrowed by atherosclerosis, resulting in an inadequate blood supply to the heart muscle [1, 2]. Narrowing of the coronary arteries can lead to sudden blockage of the coronary arteries, resulting in

myocardial infarction or even sudden cardiac death [3, 4]. The World Health Organization has identified CAD as the leading cause of death in both developing and developed countries [5, 6]. CAD is reported to be responsible for approximately 30% of annual deaths worldwide. However, more than 60% of the global burden of CAD occurs in developing countries [7]. CAD is also the leading cause of death in the United States [3]. In China, an estimated 290 million people suffer from heart disease, and the death rate from heart disease is more than 40% [7]. In Iran, cardiovascular diseases are the leading cause of death, and one million disability-adjusted life years (DALY) have been reported from this disease, of which CAD is the largest contributor [8]. Furthermore, CAD accounts for 20-23% of the disease burden in the country and cause 46% of all deaths.

Early diagnosis of CAD is very important and life-saving. However, the diagnosis of CAD is a very challenging issue. Signs and symptoms overlap with other diseases and can mislead physicians [9, 10]. In many cases, the first manifestation of CAD is myocardial infraction [11, 12]. The gold standard method for diagnosing CAD is Coronary Angiography (CA). However, CA is an expensive and invasive procedure that requires a high level of technology and technical experience [4]. Unfortunately, this diagnostic method is currently overused [13]. Although there are guidelines to avoid this invasive procedure in low-risk patients, for example, in the United States, a significant number of patients who underwent this procedure had a normal CA [14, 15]. Computed Tomography coronary Angiography (CTA) is an alternative, non-invasive, and highly accurate method of assessing CAD [16]. The European Society of Cardiology recommends that the use of CTA be considered only in patients with a pre-test probability of CAD of 15-50% [17]. However, CTA is not widely available and easily accessible method, and cannot be used in all patients, for example in pregnant women or those with an allergic reaction to contrast dye [18]. Therefore, an accessible and non-invasive method for screening patients with suspected CAD is still needed to reduce the overuse of CTA and its associated risks while increasing the accuracy of diagnosis.

Numerous studies have attempted to determine the excessive utilization of CA and its contributing factors [19-24]. However, these investigations have failed to alleviate physicians' apprehensions regarding overlooking the detection of potentially perilous CAD, prompting the continued widespread application of this diagnostic procedure. In recent years, artificial intelligence (AI) techniques have been explored for diagnosing heart disease and assessing its risks [5]. Emerging diagnostic techniques employing computer systems equipped with machine learning algorithms hold promise as non-invasive approaches for heart disease detection. Gokulnath et al. developed a classification model based on genetic algorithms and support vector machines, achieving an impressive 88.34% accuracy in distinguishing patients with CAD [25]. Similarly, Das et al.'s neural network ensemble demonstrated outstanding performance, attaining an accuracy of 89.01% in CAD diagnosis [26]. Mohan et al. presented a model for CAD detection with an accuracy of 88.4% [27]. Zhenya et al.'s hybrid cost-sensitive ensemble model further enhanced the diagnostic accuracy, producing a specificity of 93.21% [7]. Forrest et al.'s random forest model-based quantitative marker for CAD diagnosis yielded an area under the ROC curve (AUC) of 0.91, demonstrating remarkable discriminatory ability [28]. Alizadeh-Sani et al.'s sequential minimal optimization (SMO) model-powered CAD detection approach surpassed the 90% accuracy benchmark, reaching 94% [29]. These advancements in AI-driven diagnostic methods pave the way for more precise, risk-free, and cost-effective CAD evaluation for patients [7].

The integration of artificial intelligence approaches has the potential to alleviate the substantial burden and risks associated with diagnostic procedures like CA and CTA, benefiting both the healthcare system and patients. It serves as an efficient tool for screening individuals with suspected low or intermediate risk of CAD. However, to our understanding, artificial intelligence techniques have not been applied in the screening of patients suspected of CAD through routine assessments or tests. In this study, our aim is to assess and compare various machine learning techniques (including K-nearest neighbors, artificial neural network, support vector machine (SVM), decision tree, logistic regression, and stacked ensemble) in identifying CAD patients using non-invasive (routine) clinical and paraclinical data obtained from both healthy individuals and those with CAD. We were intended to identify machine learning algorithms that could effectively support physicians in diagnosing CAD in patients with low or intermediate risk, thereby contributing to a reduction in both associated risks and costs. Furthermore, the distinctive contributions of this study compared to similar research are twofold: First, to the best of our knowledge none of the prior studies in this domain have determined the importance and weight of clinical and paraclinical parameters in CAD diagnosis. In alignment with the clinical significance of this matter and in accordance with the MI-CLAIM checklist [30] recommendations, our study explores the weight and significance of the parameters influencing CAD diagnosis. Second, feature selection methods are employed to identify the minimal variables that bear the most significant impact on CAD diagnosis. This approach enhances the precision and efficiency of the diagnostic process by identifying the key variables that contribute

substantially to the accuracy of CAD diagnosis.

MATERIAL AND METHODS

Dataset description

In this study, we used the Z-Alizadeh Sani dataset. The dataset was collected from 303 patients who were referred to Shahid Rajaei Cardiovascular Research Center in Tehran between fall 2011 and winter 2012 [31]. Each patient had 54 recorded features that can be considered as indicators for the occurrence of CAD [32]. These features were divided into four categories: demographics, symptoms and examination, ECG and laboratory, and echo characteristics. Table 1 shows the names, definitions, and value ranges of these features. All patients belonged to one of the two categories of patients (with CAD) or healthy (without CAD). If a coronary artery's diameter was stenosed 50% or more, the person was classified as a patient with CAD.

Table 1: Description of Z-Alizadeh Sani dataset.

Feature type	Feature name	Range
Demographic	Age	30-86
	Weight	48-120
	Sex	Male, Female
	BMI (Body Mass Index: Kg/m^2)	18-41
	DM (Diabetes Mellitus)	Yes, No
	HTN (Hypertension)	Yes, No
	Current smoker	Yes, No
	Ex-smoker	Yes, No
	FH (Family History)	Yes, No
	Obesity	Yes (if MBI > 25), no (otherwise)
	CRF (Chronic Renal Failure)	Yes, No
	CVA (Cerebrovascular Accident)	Yes, No
	Airway disease	Yes, No
	Thyroid disease	Yes, No
	CHF (Congestive Heart Failure)	Yes, No
	DLP (Dyslipidemia)	Yes, No
Symptom and examination	BP (Blood Pressure: mmHg)	90-190
	PR (Pulse Rate)	50-110 /minute
	Edema	Yes, No
	Weak peripheral pulse	Yes, No
	Lung rales	Yes, No
	Systolic murmur	Yes, No
	Diastolic murmur	Yes, No
	Typical chest pain	Yes, No
	Dyspnea	Yes, No
	Function class	1,2,3,4
	Atypical chest pain	Yes, No
	Nonanginal chest pain	Yes, No
	Exertional CP (Exertional Chest Pain)	Yes, No
	Low Th Ang (low Threshold angina)	Yes, No
ECG	Q Wave	Yes, No
	ST Elevation	Yes, No
	ST Depression	Yes, No
	LVH (Left Ventricular Hypertrophy)	Yes, No
	Poor R progression	Yes, No
	BBB (Bundle Branch Block)	LBBB, RBBB ،No
Laboratory and echo	FBS (Fasting Blood Sugar: mg/dl)	62-400
	Cr (Creatine: mg/dl)	0.5-2.2
	TG (Triglyceride: mg/dl)	37-1050
	LDL (Low Density Lipoprotein: mg/dl)	18-232
	HDL (High Density Lipoprotein: mg/dl)	15-111
	BUN (Blood Urea Nitrogen: mg/dl)	6-52
	ESR (Erythrocyte Sedimentation Rate: mm/h)	1-90
	Hb (Hemoglobin: g/dl)	8.9-17.6
	K (Potassium: mEq/lit)	3-6.6
	Na (Sodium: mEq/lit)	128-156
	WBC (White Blood Cell: cells/ml)	3700-18000
	Lymph (Lymphocyte: %)	7-60
	Neut (Neutrophil: %)	32-89
	PLT (Platelet: 1000/ml)	25-742
	EF (Ejection Fraction: %)	15-60
	Region with RWMA (Regional Wall Motion Abnormality: number)	0,1,2,3,4
	VHD (Valvular Heart Disease)	Normal, mild, moderate, severe

Preprocessing

Data preprocessing plays a crucial role in converting raw data into a format compatible with machine learning algorithms, thereby enhancing their performance. This process involves employing diverse techniques like handling missing values, eliminating outliers, discretization, scaling, and more. It is worth noting that the dataset under consideration did not contain any missing values. In this investigation, we applied discretization and standardization approaches to enhance the efficacy of machine learning algorithms, as elucidated in the subsequent sections.

Standardization

To eliminate the undue effects of different scales on the algorithm performance, the standardization method was used. It ensured that the mean and variance of the continuous features were equal to zero and one, respectively. The standardization formula shown in (1) is used for this purpose.

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Where X represents the value, we intend to standardize, μ denotes the average of the samples, and σ represents the standard deviation of the samples.

Discretization

Discretization of continuous variables (features) can improve classification performance [33]. In this study, the discretization method was utilized to discretize the values of certain continuous features, such as the age feature, into three categories. Table 2 shows the remaining continuous features that were discretized according to Braunwald's heart book [32].

Table 2: Discritized features and their range of values

Feature name	Discretization and range of values		
Age	18-35	36-55	55<
FBS	70>	70-105	105<
LDL	≥130		130<
HDL	35>		≤35
BUN	7>	7-20	20<
ESR	If male and ESR ≤ age/2 or if female and ESR ≤ age/2 + 5		If male and ESR > age/2 or if female and ESR > age/2 + 5
Hb	If male and Hb < 14 Or If female and Hb < 12.5	If male and 14 ≤ Hb ≤ 17 or if female and 12.5 ≤ Hb ≤ 15	If male and Hb > 17 or if female and Hb > 15
K	3.8>	3.8-5.6	5.6<
Na	136>	136-146	146<
WBC	4000>	4000-11000	11000<
PLT	150>	150-450	450<
BP	90>	90-140	140<
PulseRate	60>	60-100	100<
TG	≥200		200<

Feature selection

The primary aim of this study was to differentiate CAD using routine and readily available tests. To align with this objective, the initial step involved the exclusion of three echo-related features. Feature selection in machine learning entails the process of identifying and choosing a subset of the most relevant features from a dataset for the construction of machine learning models. The utilization of feature selection algorithms serves to streamline models, enhance interpretability, decrease training time, and ultimately improve overall model performance. In the context of this specific research, two distinct feature selection techniques, namely LASSO and ReliefF, were implemented to refine the selection of features for subsequent analysis and model development.

LASSO feature selection algorithm

The LASSO algorithm chooses features by updating the absolute value of the feature coefficient in the regression model [34]. As the coefficients are updated, features with zero coefficients are eliminated from the subset. The algorithm selects features with significant coefficients, discarding others, thereby simplifying the machine learning model and enhancing interpretability.

ReliefF feature selection algorithm

The ReliefF algorithm employs instance-based learning, assigning weights to features based on their importance [35]. Feature weights indicate their ability to differentiate between classes (healthy vs. CAD patients). Features are ranked according to their weights, and those exceeding a predefined threshold are chosen for the final subset. ReliefF selects crucial features with the most impact on the target class. The algorithm operates by randomly selecting samples from the training set and determining the closest samples of the same and opposite classes for each. Feature weights are updated based on how effectively their values distinguish between the chosen sample and its nearest neighbors of the same and opposite classes. Higher weights signify greater feature importance.

Machine learning algorithms

In this research, six widely recognized classifiers, SVM, LR, KNN, ANN, DT, and SE were constructed to differentiate between patients with CAD and healthy individuals. Notably, ensemble learning models like SE have been relatively underutilized in prior studies within the CAD domain, while SVM, ANN, DT, and KNN models have been more frequently employed [36].

Support vector machine (SVM)

The SVM is a powerful machine learning model. This model creates a hyperplane, or set of hyperplanes, in a high-dimensional space that can be used for classification, regression, or other tasks. Good separation is achieved by a hyperplane with the largest distance to the nearest training data points of each class, because in general, the larger the distance, the lower the generalization error of the model. SVMs are particularly useful for classifying small or moderately complex data sets [37, 38].

Logistic regression (LR)

LR, also known as logit regression, is a widely used method for predicting the probability that a sample belongs to a particular class. If the probability is

greater than 50%, the model classifies it as belonging to the positive class (labeled one); otherwise, it predicts it as belonging to the negative class (labeled zero). This property makes logistic regression an effective binary classifier [39].

K-nearest neighbor (KNN)

The KNN is a supervised learning algorithm that classifies an instance based on the labels of its k-nearest neighbors in the feature space. Neighbors are identified using a distance function, which is chosen based on the types of features in the dataset. The label of a sample is defined by the highest number of labels corresponding to its k-nearest neighbors [40].

Artificial neural network (ANN)

The multilayer perceptron (MLP) is a supervised learning algorithm and a fully connected class of feedforward artificial neural networks. MLPs can be used for both classification and regression tasks. An MLP consists of at least three layers: an input layer, a hidden layer, and an output layer [37].

Decision tree (DT)

The decision tree algorithm is a powerful machine learning algorithm that can perform classification, regression, and even multi-output tasks. In this study, we used the classification and regression tree (CART) algorithm to train the decision tree, also known as the "growing" tree. This algorithm works by first dividing the training set into two subsets using a feature k and a threshold tk. It searches for the pair (k, tk) that produces the purest subsets. Once the CART algorithm has successfully split the training set into two parts, it continues to split the subsets and sub-subsets using the same logic. The process stops when the maximum depth is reached, defined by the max_depth hyperparameter, or when no partition can be found that reduces the impurity [37].

Stacked ensemble learning (SE)

A group of predictors is called an ensemble, and an ensemble learning algorithm is called an ensemble method. Stacked generalization is one of the ensemble learning methods that works by inferring the biases of the generalizer(s) with respect to a given training set. In this method, each of the base predictors predicts a different value, and then the final predictor (called the blender or meta learner) takes these predictions as inputs and makes the final prediction [37, 41]. In the present study, SVM, DT, ANN, and KNN models were selected as the basic classifiers, and the LR model was selected as the meta-learner. The hyperparameters of the developed models, which were selected using the grid search method, are presented in Table 3. The hyperparameters presented in the mentioned table were used for the basic classifiers in the SE models.

Table 3: Hyperparameters of models developed using LASSO and ReliefF feature selection method.

Feature selection method	Model	Hyperparameter		
LASSO	KNN	weights = 'distance'	p=1	n_neighbors=12
	SVM		kernel='sigmoid'	C=80
	ANN	max_iter =1000	hidden_layer_sizes = (2)	activation='logistic'
	LR	'penalty': 'l2'	solver='sag'	C=1
	DT	splitter='random'	max_depth=8	criterion='entropy'
ReliefF	KNN	weights='uniform'	p=1	n_neighbors=5
	SVM		kernel='linear'	C=1
	ANN	max_iter =1000	hidden_layer_sizes=(3,)	activation='tanh'
	LR	'penalty': 'l2'	solver='lbfgs'	C=1
	DT	splitter='random'	max_depth=5	criterion='gini'

Experimental Setup

This experiment was implemented in the Colab Jupyter notebook environment using Python 3.10.11 and applying the following libraries: Scikit-learn 1.2.2, NumPy 1.22.4, Pandas 1.5.3, and SHAP 0.41.0.

Machine learning models evaluation

We utilized the MI-CLAIM checklist for result reporting [30]. Performance evaluation of the learning models was conducted using the confusion matrix, with metrics including accuracy, precision (or positive predictive value (PPV)), F1 score, sensitivity, specificity, negative predictive value (NPV), and AUC (area under the curve (ROC)). In this investigation, the two classes are class one (CAD positive - individuals with CAD) and class zero (CAD negative - healthy individuals). Correct classification into the positive class is termed true positive (TP), while correct classification into the negative class is termed true negatives (TN). Misclassifying samples from the positive class as negative is denoted as false negatives (FN), and misclassifying samples from the negative class as positive is termed false positives (FP). These frequencies enable the calculation of classification

performance metrics, reflecting the classifier's effectiveness in detecting the specified class. Commonly used evaluation metrics include recall or sensitivity (2), F1 score (3), accuracy (4), precision (5), specificity (6), and AUC. Notably, for screening a severe condition like CAD, a high sensitivity metric holds greater importance in evaluating the CAD screening model than the specificity metric.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (4)$$

$$\text{Precision (PPV)} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

In this research, we employed stratified K-fold cross-validation to assess the models' performance. This technique involves the random division of the dataset into K partitions. For instance, with K set at 10, the model undergoes 10 iterations. During each iteration, one partition is designated for validation, while the remaining K-1 partitions serve as training data. Subsequently, the average of the K results is computed. Past studies across different datasets and machine learning approaches have indicated that 10 is a suitable value for error estimation and model evaluation [42].

Model explaining

While machine learning models have notably enhanced the predictive capabilities for diseases, length of stay, complications, and future patient outcomes, a substantial drawback lies in the challenge of interpreting outputs from highly complex models. This limitation hinders the effective use of these models in clinical settings, as clinicians require insight into the factors influencing a prediction for targeted interventions. Consequently, clinicians often prefer more interpretable and simpler models, such as linear models, even if they exhibit lower accuracy, over complex yet more accurate models [43].

To address this concern, our study employed the SHAP (SHAPley Additive exPlanations) technique to elucidate the output of the ANN model. SHAP is a method rooted in game theory, offering explanations and interpretations for machine learning models. By utilizing SHAPley values, it reveals the significance and contribution of each feature to the model's output [43]. In our study, we leveraged 227 samples as background data and computed SHAPley values for 76 samples to enhance the interpretability of the model.

RESULTS

The dataset consisted of 303 samples with an average age of 58.9 years, comprising 127 women and 176 men. After pre-processing the data, the LASSO and ReliefF feature selection techniques were employed to identify the most important features from the initial 51 (excluding 3 echo-related features). The 32 selected features, ordered by their importance, are presented in Table 4. Subsequently, six machine learning classification models—ANN, SVM, KNN, LR, DT, and SE—were developed using different subsets of these 32 features. The best results for each of the six models were reported based on the stratified 10-fold cross-validation technique.

Table 5 presents the evaluation results of models utilizing the top 25 selected features with the LASSO technique. In terms of accuracy, the proposed SE model achieved the highest performance, reaching an impressive 90.39%. Following the SE model, the ANN and SVM models, employing the top 22 selected features, demonstrated the next highest accuracies. Specifically, the ANN model achieved an accuracy of 90.38%, while the SVM model attained an accuracy of 90.07%. Notably, the ANN model exhibited the highest sensitivity at 94.43%, surpassing the SE and LR models with sensitivities of 93.96% and 93.52%, respectively.

Regarding specificity, the DT model led with the highest value at 83.75%, while the SVM and KNN models displayed specificities of 82.77% and 82.63%, respectively. In terms of the F1-score metric, the ANN model outperformed others with the highest score at 93.36%. The SVM model demonstrated the highest AUC at 93.22%, followed by the ANN and SE models with AUCs of 92.46% and 92.21%, respectively. Additionally, the ANN model achieved the highest NPV at 85.95%, with the SE and SVM models recording NPVs of 85.35% and 83.94%, respectively. For precision, the DT model exhibited the highest value at 93.57%, while the SVM and KNN models displayed precisions of 93.28% and 92.98%, respectively.

Following the exploration of various subsets of selected features using the ReliefF feature selection technique to develop machine learning models, the results were detailed in Table 6. The SE model, employing the top 27 selected features, demonstrated the highest accuracy at 90.06%. Subsequently, the ANN and SVM models, utilizing the top 18 selected features, achieved accuracies of 88.79% and 88.77%, respectively. Notably, the SE model exhibited the highest sensitivity among the models, reaching 93.98%. Following closely, the ANN model recorded a sensitivity of 92.12%, while the LR model displayed a sensitivity of 91.68%. Specificity was highest for the DT model at 82.36%. The SVM and KNN models achieved the highest specificities following the DT model, with values of 81.38% and 80.27%, respectively.

In terms of the F1-score metric, the SE, ANN, and SVM models led with the highest scores, achieving percentages of 93.12%, 92.13%, and 92.05%, respectively. Based on the AUC metric, the ANN model secured the highest AUC at 92.24%, with the SVM and SE models obtaining the next highest AUC values of 91.71% and 91.30%, respectively. The SE model also achieved the highest NPV at 85.33%, followed by both the ANN and SVM models at 81.33%. For precision, the DT model led with the highest value at 93.27%, while the SVM and SE models displayed precision metrics equal to 92.74% and 92.48%, respectively.

Table 4: Evaluating the importance of the variables (features) using the LASSO and ReliefF algorithms.

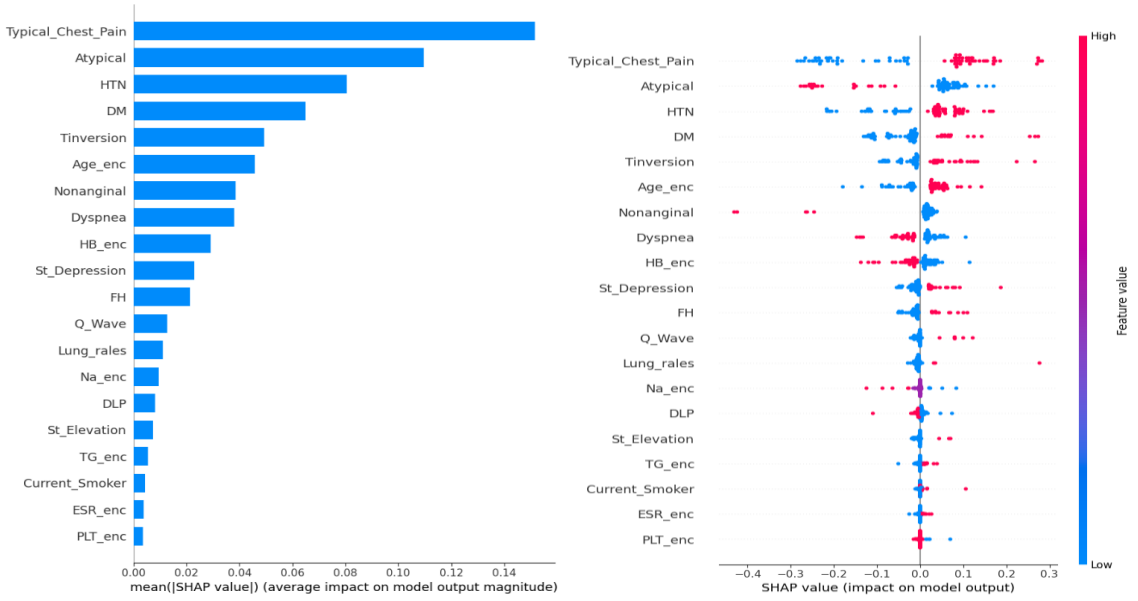
Rank	Selected feature	Coefficient	Rank	Selected feature	Weight
1	'Typical_Chest_Pain'	2.21	1	Typical_Chest_Pain	0.322772
2	'DM'	1.59	2	Atypical	0.231023
3	'Q_Wave'	1.49	3	Age_enc	0.122112
4	'Tinversion'	1.31	4	HTN	0.119802
5	'Nonanginal'	-1.3	5	DM	0.082838
6	'Age_enc'	1.27	6	Nonanginal	0.058086
7	'HTN'	1.14	7	Tinversion	0.051815
8	'FH'	1.08	8	FBS_enc	0.048845
9	'PLT_enc'	-0.93	9	BP_enc	0.048185
10	'Dyspnea'	-0.81	10	Current_Smoker	0.042574
11	'Lung_rales'	0.72	11	Dyspnea	0.036304
12	'TG_enc'	0.58	12	TG_enc	0.033003
13	'ST_Depression'	0.57	13	Weight_stndrd	0.032765
14	'Atypical'	-0.56	14	BMI_stndrd	0.029858
15	'DLP'	-0.55	15	Sex	0.029703
16	'Na_enc'	-0.43	16	Neut_stndrd	0.026767
17	'CR_enc'	-0.4	17	DLP	0.022772
18	'St_Elevation'	0.39	18	Systolic_Murmur	0.020132
19	'HB_enc'	-0.38	19	Length_stndrd	0.018461
20	'Current_Smoker'	0.34	20	BUN_enc	0.015842
21	'PR_enc'	-0.33	21	LVH	0.013201
22	'ESR_enc'	-0.27	22	Lymph_stndrd	0.012193
23	'Neut_stndrd'	0.27	23	Edema	0.010231
24	'BP_enc'	0.21	24	BBB	0.010231
25	'Systolic_Murmur'	0.18	25	LDL_enc	0.008911
26	'Function_Class'	0.12	26	FH	0.006931
27	'BBB'	-0.12	27	Function_Class	0.006601
28	'HDL_enc'	-0.11	28	ST_Elevation	0.005281
29	'BMI_stndrd'	-0.11	29	PLT_enc	0.00495
30	'Sex'	0.08	30	HB_enc	0.00495
31	'Diastolic_Murmur'	-0.07	31	Thyroid_Disease	0.00462
32	'LDL_enc'	-0.04	32	Diastolic_Murmur	0.00462

Table 5: Evaluation metrics of the models developed using the LASSO feature selection technique.

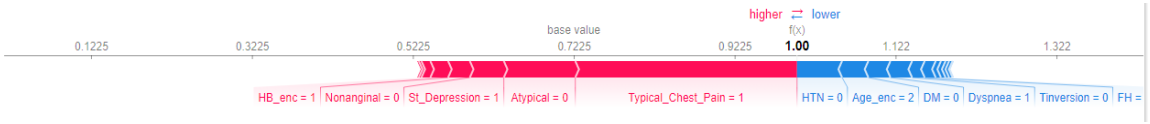
Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC	NPV	Top used features
KNN	0.8812	0.9030	0.8263	0.9298	0.9150	0.8855	0.7835	10
SVM	0.9007	0.9303	0.8277	0.9328	0.9302	0.9322	0.8394	22
ANN	0.9038	0.9443	0.8027	0.9251	0.9336	0.9246	0.8595	22
LR	0.8873	0.9352	0.7680	0.9132	0.9228	0.9217	0.8282	24
SE	0.9039	0.9396	0.8152	0.9291	0.9332	0.9221	0.8535	25
DT	0.8911	0.9123	0.8375	0.9357	0.9223	88.28	0.8058	11

Table 6: Evaluation metrics of the models developed using the ReliefF feature selection technique.

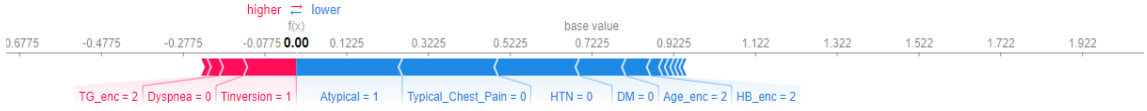
Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC	NPV	Top used features
KNN	0.8845	0.9166	0.8027	0.9224	0.9189	0.9028	0.7982	13
SVM	0.8877	0.9166	0.8138	0.9274	0.9205	0.9171	0.8133	18
ANN	0.8879	0.9212	0.8013	0.9229	0.9213	0.9224	0.8133	18
LR	0.8810	0.9168	0.7930	0.9201	0.9164	0.9111	0.8091	19
SE	0.9006	0.9398	0.8013	0.9248	0.9312	0.9130	0.8533	27
DT	0.8810	0.9030	0.8236	0.9327	0.9154	0.8848	0.7849	7



(a) **(b)**



(c)



(d)

Fig 1: SHAP summary plot for the top 20 features that contributing to the ANN model

The SHAP technique

In the context of feature selection through the LASSO technique, wherein a subset of 22 features was identified, the ANN model demonstrated superior sensitivity. Subsequent to this selection, an analysis of the importance and contribution of each feature to the ANN model was conducted using SHAP values. In Fig 1(a), the presentation encompasses the top 20 features of significance, delineated by the average absolute SHAP value. Within this figure, the assessment of feature importance and its corresponding contribution to the model's output is organized in a descending order, signifying the hierarchical ranking of features based on their respective significance. Consequently, the feature occupying the highest rank within this sequence is deemed to exert the greatest importance on the model's output.

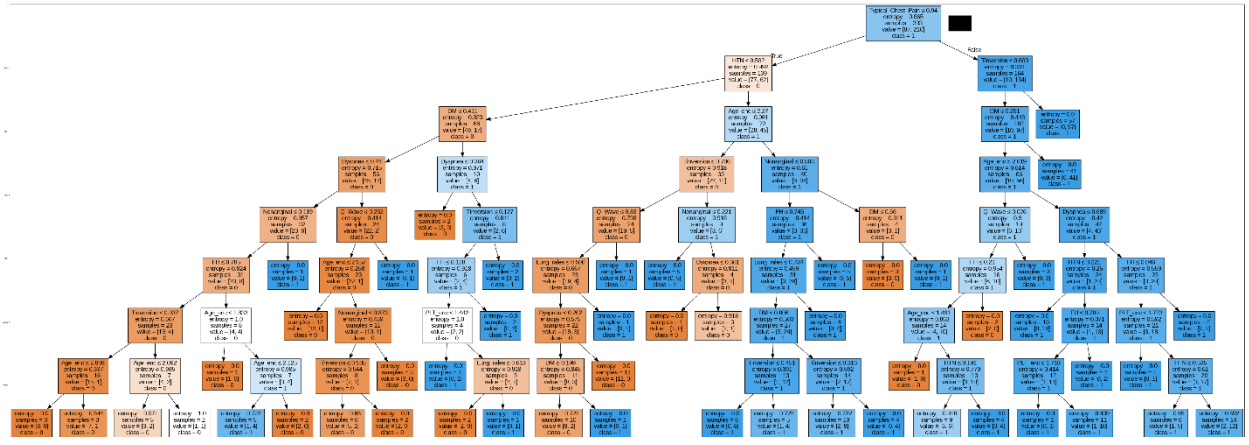
Fig 1(b) employs a color spectrum ranging from blue to red to visually represent the values associated with each feature. In this representation, the blue color denotes the lowest values, while the red color signifies the highest values of the respective features. Notably, in this visualization, the "typical chest pain" feature emerges as the most prominently colored, suggesting its heightened importance and substantial impact on the predictive outcomes of the ANN model. To elucidate, if an individual exhibits characteristic such as typical chest pain, hypertension (HTN), diabetes mellitus (DM), and T inversion, these features collectively exert a positive influence, thereby contributing to a model prediction favoring classification into the CAD class, and conversely for those lacking these attributes.

Fig 1(c) and 1(d) provide detailed insights into the significance and contribution of individual features within the output of the ANN model for two randomly selected cases. In Fig 1(c), the model has categorized the case as belonging to the CAD class based on the values of the features. Notably, the presence of typical chest pain and ST-depression, coupled with the absence of atypical chest pain, emerged as features with substantial contributions to the predictive outcome of this model.

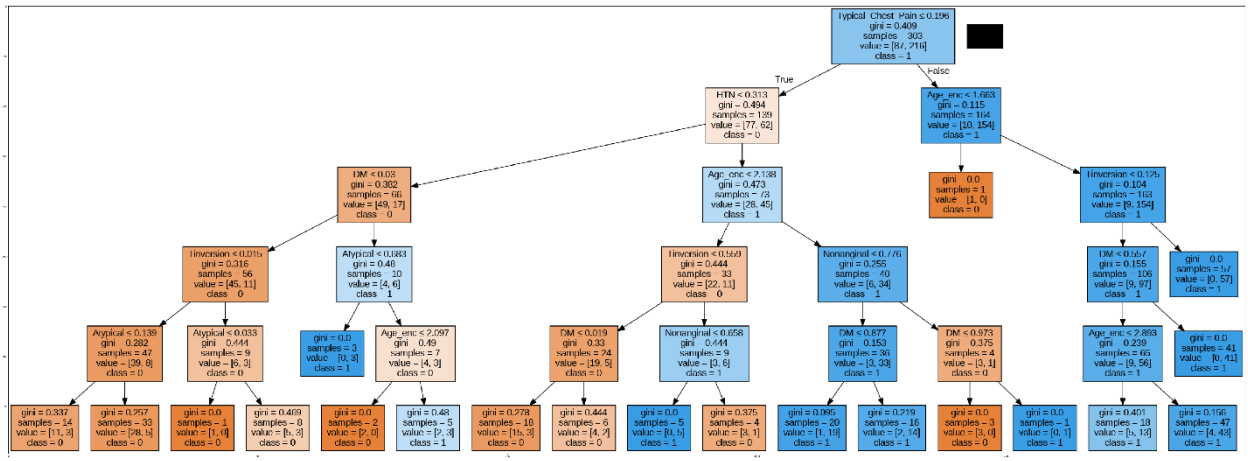
Fig 1(d) further illustrates the importance and contribution of features for a case identified by the model as belonging to the healthy class. In this instance, the absence of typical chest pain, hypertension (HTN), and diabetes mellitus (DM), along with the presence of the "Atypical chest pain" feature, exerted the most substantial impact on the model's output. This collective evidence suggests that these specific features were influential indicators that led the model to predict the absence of CAD in this case.

Classification And Regression Tree (CART)

In this study, decision trees were generated using two distinct feature selection techniques—LASSO and ReliefF. Both techniques demonstrated favorable outcomes in terms of accuracy, sensitivity, and specificity metrics. Notably, the decision tree constructed with the top 11 features selected through the LASSO technique (depicted in Fig 2(a)) exhibited slightly superior performance compared to the decision tree formulated with the top 7 features selected by the ReliefF technique (illustrated in Fig 2(b)). However, it is noteworthy that the decision tree derived from the ReliefF feature selection algorithm, with a depth of 5, showcased greater simplicity and comprehensibility when contrasted with the decision tree developed through the LASSO algorithm, which had a depth of 8.



(a): Top 11 features selected by the LASSO feature selection method



(b): Top 7 features selected by the ReliefF feature selection method.

Fig 2: Decision trees generated using selected features.

DISCUSSION

In this study, we employed two feature selection algorithms, namely LASSO and ReliefF, to identify the most pivotal features within the Z-Alizadeh Sani dataset. The LASSO algorithm identified the top 10 features, ranked in order of significance, as follows: Typical chest pain, diabetes mellitus, Q wave, T-inversion, non-anginal chest pain, age, hypertension, family history, platelet count, and dyspnea. Similarly, the ReliefF algorithm selected the top 10 features out of the initial 32, arranged by their importance: Typical chest pain, atypical chest pain, age, hypertension, diabetes mellitus, non-anginal chest pain, T-inversion, fasting blood sugar, blood pressure, and current smoker. Utilizing various subsets of these selected features, we developed and assessed six machine learning models—DT, LR, KNN, ANN, SVM, and SE. Evaluation was performed using a 10-fold cross-validation approach. Overall, all models exhibited robust performance, as evidenced by the high metrics presented in Tables 5 and 6. These findings underscore the considerable potential of machine learning algorithms in effectively discerning CAD.

The ANN, SVM, and SE models demonstrated higher performance using both feature selection methods. However, when comparing the performance of the models across the feature selection methods, the performance of the ANN, SVM, and SE models was slightly better using the features selected by the LASSO method. A potentially effective screening AI model should use fewer but more accessible features to develop a model while maintaining high performance. The most important performance indicator for a screening method is the sensitivity test. Because the method should not miss any true case of CAD. Therefore, the ANN model developed using LASSO feature selection is a better candidate for our purpose. This model also showed better performance when considering other performance indicators. Furthermore, the ANN model is preferable to the SE model in the LASSO group because of the smaller number of features used by the model.

CAD is the leading cause of death in both developing and developed countries. Correct and timely diagnosis of this disease is very important but also very challenging. This is because the signs and symptoms of this disease are complex, and in some cases they may be asymptomatic, or the symptoms may overlap with those of other diseases. The accurate methods of diagnosing CAD are expensive, inaccessible to the broader population, and associated with potential side effects. Therefore, it can be very helpful to use a more accessible and cost-effective screening method before proceeding with these advanced methods.

The ANN model, commonly referred to as a black box, posed a challenge in terms of interpretability. To address this, the SHAP model explanation technique was employed to elucidate and quantify the importance and contribution of features within the ANN model, which demonstrated high performance utilizing 22 features. The utilization of the SHAP technique is deemed valuable as it offers clinicians a more comprehensive understanding of the model. The SHAP results highlighted that typical chest pain, atypical chest pain, hypertension, diabetes mellitus, T inversion, and age were the most impactful features influencing the output of the ANN model. In addition to the ANN model, DT models were developed in this study due to their simplicity and ease of interpretation. The decision tree derived from the ReliefF feature selection algorithm, with a depth of 5, exhibited greater simplicity and comprehensibility compared to the decision tree developed through the LASSO algorithm, which had a depth of 8.

Recently, machine learning techniques have garnered

significant attention for the development of robust tools in predicting CAD. Numerous studies in this domain have yielded promising outcomes. Alizadeh Sani et al. [29] conducted a study wherein the SMO model, featuring 36 SVM-selected features, achieved notable performance with 93.39% accuracy, 95.37% sensitivity, and 88.51% specificity, outperforming NB, SMO bagging, and ANN models. In a related study, our ANN model, utilizing LASSO feature selection, achieved a sensitivity of 0.9443 with a smaller set of routine features, surpassing the performance of the ANN model in the aforementioned study. Arabasadi et al. [44] introduced a hybrid neural network-genetic algorithm model, demonstrating a sensitivity of 97%, specificity of 92%, and an accuracy of 93.85%. Hassannataj et al. [45] illustrated improved SVM performance using the ANOVA kernel combined with GA, achieving 89.45% accuracy, 81.22% sensitivity, 100% specificity, 100% PPV, and 92.9% NPV. Velusamy et al. [46] utilized an ensemble voting technique based on weighted-average voting (WAVEn), achieving 98.97% accuracy, 100% sensitivity, and 96.30% specificity. Notably, the top features in these studies often derived from echocardiography, an accessibility challenge for the general population.

Joloudari et al. [47] demonstrated superior performance of the random trees (RT) model, using 40 features, compared to SVM, C5.0 decision tree, and CHAID decision tree models. Our DT and SVM models, developed with LASSO and ReliefF, outperformed these models in accuracy and AUC metrics with fewer features. Ghiasi et al. [48] developed a CART decision tree model with five features, achieving 92.41% accuracy, 77.01% specificity, and 98.61% sensitivity. While our DT models using LASSO and ReliefF feature selection algorithms showed comparable accuracy and sensitivity, they offer simplicity, interpretability, and do not rely on echocardiography features. Dahal et al. [49] emphasized the superiority of the SVM model in diagnosing CAD over LR, KNN, Bagging CART, and RF models, achieving an accuracy of 89.47%, sensitivity of 94.34%, specificity of 78.26%, and an AUC of 88.68%. Table 7 provides a comparative overview of studies conducted on the Z-Alizadeh Sani dataset.

Table 7: A compilation of previous studies conducted on the Z-Alizadeh Sani dataset.

Algorithms + feature selection technique	Accuracy	Sensitivity	Specificity	NPV	AUC	Number of selected features
SMO + weight by SVM Alizadehsani et al. [29]	93.39	95.37	88.51	_	_	34
ANN + weight by SVM Alizadehsani et al. [29]	87.13	90.28	79.31	_	_	34
ANN + t-test + PCA Cüvitoğlu et al. [50]	85.15	72.36	90.26	_	93	25
CART Ghiasi et al. [48]	92.41	98.61	77.01	_	_	5
SVM + Chi-square Dahal et al. [49]	89.47	94.34	78.26	_	88.68	15
LR + Chi-square Dahal et al. [49]	86.84	94.34	69.57	_	90.32	15
KNN + Chi-square Dahal et al. [49]	71.05	92.45	21.74	_	58.94	15
WAVEn + Boruta wrapper + SVM Velusamy et al. [46]	98.97	100	96.30	_	_	5
Random trees Joloudari et al. [47]	91.47	_	_	_	96.70	40
SVM + GA Hassannataj et al. [45]	89.45	81.22	100	92.90	100	31

Machine learning algorithms have demonstrated remarkable performance in this area, but the interpretability of these complex models remains a formidable challenge [51]. Despite the superior performance of complex models in many scenarios, their interpretability is of paramount importance in settings such as healthcare, where conscientious and responsible decision making is imperative. The attractiveness of simpler and more understandable models is particularly evident in clinical settings. Recognizing this importance, we used the SHAP model explanation technique in our study to discern the importance and contribution of each feature and its impact on the output of the developed ANN model. Angina or chest pain stands out as one of the most common symptoms of CAD. The typical type proves to be more diagnostic than the atypical type, as the latter may be associated with various causes such as

digestive disorders and pericarditis [52, 53]. Consistent with these observations, the SHAP technique in our study identified typical chest pain as the most important feature for CAD diagnosis using the ANN model. Numerous studies have consistently highlighted hypertension, diabetes mellitus, and age as key factors in CAD [52, 54, 55]. In our investigation, hypertension, diabetes mellitus, and age emerged as the most influential features affecting the outcome of the ANN model. It's important to note that the SHAP model explanation technique, like other model explanation techniques, has its limitations [56, 57]. In particular, when applied to machine learning models that lack inherent causality, SHAP cannot accurately address causal issues.

Our study is subject to several limitations. Firstly, the dataset employed in this study had a restricted number of samples sourced exclusively from a single medical center, compromising the generalizability of the machine learning models. Additionally, this study adopts a retrospective design, relying on historical data collection. Despite the favorable results obtained by the machine learning models, there is a need for prospective studies to comprehensively assess and validate these models. Furthermore, the Z-Alizadeh Sani dataset featured an unequal distribution of patients and healthy individuals. This imbalance introduces the potential for bias, thereby influencing the outcomes and interpretation of the models. Addressing these limitations through larger and more diverse datasets, prospective study designs, and efforts to balance sample distributions would enhance the robustness and applicability of the findings.

CONCLUSION

The timely diagnosis of CAD is a critical imperative to prevent mortality and severe complications, yet it remains a formidable challenge. In this study, we addressed this challenge by developing machine learning models utilizing routine clinical and laboratory findings to identify individuals at high risk of CAD. The comprehensive set of machine learning models employed, including KNN, SVM, LR, ANN, DT, and SE, aimed to enhance screening accuracy. Notably, the ANN model, developed using the LASSO feature selection method, demonstrated high performance and emerged as a robust candidate for this screening purpose. However, it is crucial to emphasize the need for further assessments regarding external validity and generalizability before deploying the model for widespread screening in clinical environments.

AUTHOR'S CONTRIBUTION

HP and SN designed the study. SN prepared the dataset, developed the models, and performed the early data analysis using Python. HP supervised the process of model development and data analysis, while KK and ZN commented the process. SN prepared the early draft of the paper. KK and ZN commented on the draft and HP finalized the submitting version. .

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

ETHICS APPROVAL

This study was derived from the MSc. thesis of the first author under the supervision of the last author. The research ethics committee of Urmia University of Medical Sciences has reviewed and approved the study proposal with the granting of ethical code of IR.UMSU.REC.1402.082.

REFERENCES

1. Marateb HR, Goudarzi S. A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. J Res Med Sci. 2015; 20(3): 214-23. PMID: 26109965 [PubMed]
2. Gopalan C, Kirk E. Coronary artery disease. In: Gopalan C, Kirk E. Biology of cardiovascular and metabolic diseases. Academic Press; 2022.
3. Ghadiri Hedeshi N, Saniee Abadeh M. Coronary artery disease detection using a fuzzy-boosting PSO approach. Comput Intell Neurosci. 2014; 2014: 783734. PMID: 24817883 DOI: 10.1155/2014/783734 [PubMed]
4. Zeitouni M, Clare RM, Chiswell K, Abdulrahim J, Shah N, Pagidipati NP, et al. Risk factor burden and long-term prognosis of patients with premature coronary artery disease. J Am Heart Assoc. 2020; 9(24): e017712. PMID: 33287625 DOI: 10.1161/JAHA.120.017712 [PubMed]
5. Kim JK, Kang S. Neural network-based coronary heart disease risk prediction using feature correlation analysis. J Healthc Eng. 2017; 2017: 2780501. PMID: 29076332 DOI: 10.1155/2017/2780501 [PubMed]
6. Ali L, Rahman A, Khan A, Zhou M, Javeed A, Khan JA. An automated diagnostic system for heart disease prediction based on chi-squared statistical model and optimally configured deep neural network. IEEE Access. 2019; 7: 34938-45.
7. Zhenya Q, Zhang Z. A hybrid cost-sensitive ensemble for heart disease prediction. BMC Med Inform Decis Mak. 2021; 21(1): 73. PMID: 33632225 DOI:

10.1186/s12911-021-01436-7 [PubMed]

8. Sarrafzadegan N, Mohammmadifard N. Cardiovascular disease in Iran in the last 40 years: Prevalence, mortality, morbidity, challenges and strategies for cardiovascular prevention. Arch Iran Med. 2019; 22(4): 204-10. PMID: 31126179 [PubMed]

9. Fox KAA, Metra M, Morais J, Atar D. The myth of 'stable' coronary artery disease. Nat Rev Cardiol. 2020; 17(1): 9-21. PMID: 31358978 DOI: 10.1038/s41569-019-0233-y [PubMed]

10. Lahsasna A, Ainon RN, Zainuddin R, Bulgiba A. Design of a fuzzy-based decision support system for coronary heart disease diagnosis. J Med Syst. 2012; 36(5): 3293-306. PMID: 22252606 DOI: 10.1007/s10916-012-9821-7 [PubMed]

11. Corrado D, Thiene G, Pennelli N. Sudden death as the first manifestation of coronary artery disease in young people (less than or equal to 35 years). Eur Heart J. 1988; 9(Suppl N): 139-44. PMID: 3246245 DOI: 10.1093/eurheartj/9.suppl_n.139 [PubMed]

12. Mehta D, Curwin J, Gomes JA, Fuster V. Sudden death in coronary artery disease: Acute ischemia versus myocardial substrate. Circulation. 1997; 96(9): 3215-23. PMID: 9386195 DOI: 10.1161/01.cir.96.9.3215 [PubMed]

13. Huang X, Rosenthal MB. Overuse of cardiovascular services: Evidence, causes, and opportunities for reform. Circulation. 2015; 132(3): 205-14. PMID: 26195488 DOI: 10.1161/CIRCULATIONAHA.114.012668 [PubMed]

14. Bradley SM, Maddox TM, Stanislawski MA, O'Donnell CI, Grunwald GK, Tsai TT, et al. Normal coronary rates for elective angiography in the veterans affairs healthcare system: Insights from the VA CART program (veterans affairs clinical assessment reporting and tracking). J Am Coll Cardiol. 2014; 63(5): 417-26. PMID: 24184244 DOI: 10.1016/j.jacc.2013.09.055 [PubMed]

15. Thomas MP, Gurm HS, Nallamothu BK. When is it right to be wrong? J Am Coll Cardiol. 2014; 63(5): 427-9. PMID: 24184239 DOI: 10.1016/j.jacc.2013.10.010 [PubMed]

16. Joshi H, Shah R, Prajapati J, Bhangdiya V, Shah J, Kandre Y, et al. Diagnostic accuracy of computed tomography angiography as compared to conventional angiography in patients undergoing noncoronary cardiac surgery. Heart Views. 2016; 17(3): 88-91. PMID: 27867455 DOI: 10.4103/1995-705X.192555 [PubMed]

17. Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, Budaj A, et al. 2013 ESC guidelines on the management of stable coronary artery disease: The task force on the management of stable coronary artery disease of the European society of cardiology. Eur Heart J. 2013; 34(38): 2949-3003. PMID: 23996286 DOI: 10.1093/eurheartj/eht296 [PubMed]

18. Abbara S, Arbab-Zadeh A, Callister TQ, Desai MY, Mamuya W, Thomson L, et al. SCCT guidelines for performance of coronary computed tomographic angiography: A report of the Society of cardiovascular computed tomography guidelines committee. J Cardiovasc Comput Tomogr. 2009; 3(3): 190-204. PMID: 19409872 DOI: 10.1016/j.jcct.2009.03.004 [PubMed]

19. Chan PS, Patel MR, Klein LW, Krone RJ, Dehmer GJ, Kennedy K, et al. Appropriateness of percutaneous coronary intervention. JAMA. 2011; 306(1): 53-61. PMID: 21730241 DOI: 10.1001/jama.2011.916 [PubMed]

20. Ko DT, Guo H, Wijeysundera HC, Natarajan MK, Nagpal AD, Feindel CM, et al. Assessing the association of appropriateness of coronary revascularization and clinical outcomes for patients with stable coronary artery disease. J Am Coll Cardiol. 2012; 60(19): 1876-84. PMID: 23062534 DOI: 10.1016/j.jacc.2012.06.056 [PubMed]

21. Korenstein D, Falk R, Howell EA, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. Arch Intern Med. 2012; 172(2): 171-8. PMID: 22271125 DOI: 10.1001/archinternmed.2011.772 [PubMed]

22. Lipitz-Snyderman A, Bach PB. Overuse of health care services: When less is more ... more or less. JAMA Intern Med. 2013; 173(14): 1277-8. PMID: 23712254 DOI: 10.1001/jamainternmed.2013.6181 [PubMed]

23. Joynt KE. Rightsizing invasive cardiac services in the United States. Circulation. 2013; 128(8): 780-2. PMID: 23877257 DOI: 10.1161/CIRCULATIONAHA.113.004718 [PubMed]

24. Carr JJ, Hendel RC, White RD, Patel MR, Wolk MJ, Bettmann MA, et al. 2013 appropriate utilization of cardiovascular imaging: A methodology for the development of joint criteria for the appropriate utilization of cardiovascular imaging by the American college of cardiology foundation and American college of radiology. J Am Coll Cardiol. 2013; 61(21): 2199-206. PMID: 23500292 DOI: 10.1016/j.jacc.2013.02.010 [PubMed]

25. Gokulnath C, Shantharajah SP. An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Computing. 2019; 22: 14777-87.

26. Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications. 2009; 36(4): 7675-80.

27. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019; 7: 81542-54.

28. Forrest I, Petrazzini B, Duffy Á, Park J, Marquez-Luna C, Jordan D, et al. Machine learning-based marker for coronary artery disease: Derivation and validation in two longitudinal cohorts. The Lancet. 2022; 401(10372): 215-25.

29. Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. A data mining approach for diagnosis of coronary artery disease. Comput Methods Programs Biomed. 2013; 111(1): 52-61. PMID: 23537611 DOI:

10.1016/j.cmpb.2013.03.004 [PubMed]

30. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: The MI-CLAIM checklist. Nat Med. 2020; 26(9): 1320-4. PMID: 32908275 DOI: 10.1038/s41591-020-1041-y [PubMed]

31. Alizadehsani R, Habibi J, Hosseini MJ, Boghrati R, Ghandeharioun A, Bahadorian B, et al. Diagnosis of coronary artery disease using data mining techniques based on symptoms and ECG features. European Journal of Scientific Research. 2012; 82(4): 542-53.

32. Braunwald E, Bonow RO. Braunwald's heart disease: A textbook of cardiovascular medicine. 9th ed. Saunders; 2012.

33. Lustgarten JL, Gopalakrishnan V, Grover H, Visweswaran S. Improving classification performance with discretization on biomedical datasets. AMIA Annu Symp Proc. 2008; 2008: 445-9. PMID: 18999186 [PubMed]

34. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B. 1996; 58(1): 267-88.

35. Urbanowicz RJ, Olson RS, Schmitt P, Meeker M, Moore JH. Benchmarking relief-based feature selection methods for bioinformatics data mining. J Biomed Inform. 2018; 85: 168-88. PMID: 30030120 DOI: 10.1016/j.jbi.2018.07.015 [PubMed]

36. Alizadehsani R, Abdar M, Roshanzamir M, Khosravi A, Kebria PM, Khozeimeh F, et al. Machine learning-based coronary artery disease diagnosis: A comprehensive review. Comput Biol Med. 2019; 111: 103346. PMID: 31288140 DOI: 10.1016/j.compbiomed.2019.103346 [PubMed]

37. Geron A. Hands-on machine learning with scikit-learn, keras, and tensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.; 2019.

38. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20(3): 273-97.

39. Larsen K, Petersen J, Budtz-Jørgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. Biometrics. 2000; 56(3): 909-14. PMID: 10985236 DOI: 10.1111/j.0006-341x.2000.00909.x [PubMed]

40. Uguroglu S, Carbonell J, Doyle M, Biederman R. Cost-sensitive risk stratification in the diagnosis of heart disease. Innovative Appications of Artificial Intelligence Conference. Association for the Advancement of Artificial Intelligence; 2012.

41. Wolpert DH. Stacked generalization. Neural Networks. 1992; 5(2): 241-59.

42. Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers Inc.; 2011.

43. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018; 2(10): 749-60. PMID: 31001455 DOI: 10.1038/s41551-018-0304-0 [PubMed]

44. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Comput Methods Programs Biomed. 2017; 141: 19-26. PMID: 28241964 DOI: 10.1016/j.cmpb.2017.01.004 [PubMed]

45. Hassannataj Joloudari J, Azizi F, Nematollahi MA, Alizadehsani R, Hassannatajjeloudari E, Nodehi I, et al. GSVMA: A genetic support vector machine ANOVA method for CAD diagnosis. Front Cardiovasc Med. 2022; 8: 760178. PMID: 35187099 DOI: 10.3389/fcvm.2021.760178 [PubMed]

46. Velusamy D, Ramasamy K. Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset. Comput Methods Programs Biomed. 2021; 198: 105770. PMID: 33027698 DOI: 10.1016/j.cmpb.2020.105770 [PubMed]

47. Joloudari JH, Joloudari EH, Saadatfar H, GhasemiGol M, Razavi SM, Mosavi A, et al. Coronary artery disease diagnosis: Ranking the significant features using a random trees model. Int J Environ Res Public Health. 2020; 17(3): 731. PMID: 31979257 DOI: 10.3390/ijerph17030731 [PubMed]

48. Ghiasi MM, Zendehboudi S, Mohsenipour AA. Decision tree-based diagnosis of coronary artery disease: CART model. Comput Methods Programs Biomed. 2020; 192: 105400. PMID: 32179311 DOI: 10.1016/j.cmpb.2020.105400 [PubMed]

49. Dahal K, Gautam Y. Argumentative comparative analysis of machine learning on coronary artery disease. Open Journal of Statistics. 2020; 10: 694-705.

50. Cüvitoğlu A, Işik Z. Classification of CAD dataset by using principal component analysis and machine learning approaches. International Conference on Electrical and Electronic Engineering. IEEE; 2018.

51. The Lancet Respiratory Medicine. Opening the black box of machine learning. Lancet Respir Med. 2018; 6(11): 801. PMID: 30343029 DOI: 10.1016/S2213-2600(18)30425-9 [PubMed]

52. Davies SW. Clinical presentation and diagnosis of coronary artery disease: Stable angina. Br Med Bull. 2001; 59: 17-27. PMID: 11756201 DOI: 10.1093/bmb/59.1.17 [PubMed]

53. Hickam DH. Chest pain or discomfort. In: Walker HK, Hall WD, Hurst JW (eds). Clinical methods: The history, physical, and laboratory examinations. Butterworth Publishers; 1990.

54. Park HW, Kang MG, Kim K, Koh JS, Park JR, Jeong YH, et al. Long-term prognosis and clinical characteristics of patients with newly diagnosed diabetes mellitus detected after first acute myocardial infarction: From KAMIR-NIH registry. Korean Circ J. 2018; 48(2): 134-47. PMID: 29441746 DOI: 10.4070/kcj.2017.0174 [PubMed]

55. Weber T, Lang I, Zweiker R, Horn S, Wenzel RR, Watschinger B, et al. Hypertension and coronary

artery disease; epidemiology, physiology, effects of treatment, and recommendations : A joint scientific statement from the Austrian society of cardiology and the Austrian society of hypertension. Wien Klin Wochenschr. 2016; 128(13-14): 467-79. PMID: 27278135 DOI: 10.1007/s00508-016-0998-5 [PubMed]

56. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling lime and SHAP: Adversarial attacks on post hoc explanation methods. Conference on AI, Ethics, and Society. ACM; 2020.

57. Molnar C. Interpretable machine learning: A guide for making black box models explainable. 2nd ed. 2022.