Int Cardiovasc Res J.2022;**16**(2):72-79.icrj.122437

International Cardiovascular Research Journal

www.ircrj.com

Comparison of Random Forest and Artificial Neural Network Models to Evaluate Diagnostic Factors in the Necessity to Perform Angiography

Parastoo Golpour [1], MSc; Mohammad Tajfard [2,3], PhD; Majid Ghayour-Mobarhan [4,5], PhD; Mohsen Moohebati [6], MD; Ali Taghipour [1], MD; Habibollah Esmaily [2], PhD; Sara Sabbaghian Tousi [1,*], MSc

[1] Department of Epidemiology and Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, IR Iran
[2] Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, IR Iran
[3] Department of Health Education and Health Promotion, Faculty of Health, Mashhad University of Medical Sciences, Mashhad, IR Iran
[4] International UNESCO Center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, IR Iran
[5] Metabolic Syndrome Research Center, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, IR Iran
[6] Cardiovascular Research Center, School of Medicine, Mashhad University of Medical Sciences, Mashhad, IR Iran

ARTICLE INFO

Article Type:
Research Article

Article History:
Received: 5 Jan 2022
Revised: 2 May 2022
Accepted: 29 May 2022

Keywords:
Machine Learning, Angiography
Random Forest
Artificial Neural Network
Coronary Artery Disease
Risk Factor

ABSTRACT

Background: Coronary Artery Disease (CAD) is the most common type of cardiovascular disorders. Despite being costly and invasive, coronary angiography is a reliable method for diagnosing CAD. Therefore, it is crucial to use non-invasive methods to screen candidates for angiography to accelerate the process of decision-making. Two powerful Machine Learning (ML) methods are Random Forest (RF) and Artificial Neural Network (ANN).

Objectives: The present study aimed to compare RF and ANN to define the most important features for positive CAD results and predict the need for angiography as a screening method.

Methods: This cross-sectional study was performed on 1128 patients referred for angiography. The data were divided into test and train sets. The models (RF and ANN) were fitted with the angiographic outcome variable (positive or negative) as the dependent variable and five features as predictors. Then, the performances of the models were compared by considering the Area Under the Rock Curve (AUC). All statistical analyses were done using the R software, version 4.1.2.

Results: Out of the 1128 patients, 752 (66.7%) had positive angiography results. The AUC values were 0.75 and 0.52 for the test data set in ANN and RF models, respectively.

Conclusion: Fasting Blood Sugar (FBS), gender, age, Body Mass Index (BMI), and smoking habit were important in predicting the results of an angiography for CAD. Applying these factors in ML approaches can be considered a screen for angiography to accelerate the process of diagnosis.

1. Background

Cardiovascular Diseases (CVDs) are the leading cause of mortality globally. In 2016, approximately 17.9 million deaths occurred due to CVDs (approximately 31% of the global deaths) (1). CVDs refer to a group of disorders that involve the heart or blood vessels, with Coronary Artery Disease (CAD) being the most common one. CAD occurs when coronary arteries become hardened and narrowed (2). One reliable method for diagnosing CAD is angiography although it is costly and invasive. This procedure is also accompanied by some potential complications such as heart attack, stroke, injury to the catheterized artery, irregular heart rhythms, allergic reactions to the dye or medications used during the procedure, kidney damage, excessive bleeding, and infection (3). Therefore, it is crucial to use non-invasive methods for screening angiography candidates.

In order to derive useful information from large amounts of medical data, powerful data analysis tools are required. Machine Learning (ML) is used to improve data analysis on huge data sets. ML is also increasingly used in healthcare in order to build models for accelerating the process of

*Corresponding author: Sara Sabbaghian Tousi, Department of Epidemiology and Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad 917791-8564, Iran. Cellphone: +98- 9151116574, Email: saratoosi84@gmail.com.

medical decision-making. ML methods aim at providing a description that separate different groups correctly. Disease diagnosis is one of the applications, in which ML methods lead to successful results. In this regard, two most powerful methods are Random Forest (RF) and Artificial Neural Network (ANN).

RF was initially introduced by Breiman in 2001 (4). It is an ensemble learning method for both classification and regression that works by constructing a variety of decision trees (5). This method is made by the random selection of training data samples. Random features are also chosen in the process. Prediction is made by averaging for regression and majority vote for classification. Each tree is first grown by sampling N randomly from the original data if the training set consists of N cases with replacement. This sample is used as the training set to grow the tree. Then, m variables are randomly selected out of all input variables (M), so that m < M. After that, splitting the node is done based on the best split on these m variables. The value of m is kept constant during the forest growing. Finally, as no pruning is applied, each tree is grown to the largest possible extent (6).

Previously, decision tree and ANN were applied for predicting CAD (7, 8). Recently, progress in ML has been focused on a powerful and freely available software package; i.e., Python, which is a programming language that integrates and analyzes data more swiftly and effectively (9). ANN model assumes importance in the field of prediction such as clinical diagnosis, since it has the capability of modelling non-linear relationships in a high-dimensional dataset. It is also able to predict a complex relationship among variables. ANN works like a biological neural network. Brain consists of numerous neurons, each of which is connected with the others. Each neuron has two parts, namely dendrite and axon. The dendrite plays the role of the receiver and the axon plays the role of the information transmitter. The information is stored in the nucleus of the neuron that is transferred. Each ANN has three layers (input, hidden, and output) for receiving, processing, and reporting information (10).

2. Objectives

The present study aims to compared RF and ANN to define the most important features for predicting the need for angiography and positive CAD results.

3. Methods

3.1. Data Collection

This cross-sectional study was performed on 1187 patients referred to Ghaem Hospital, Mashhad for angiography between 2011 and 2012. Positive coronary angiography usually indicates the obstruction of more than 50% of at least one coronary artery, while negative angiography represents the obstruction of less than 50% of coronary arteries. In the data set used in the present study, patients over 18 years of age were candidates for angiography based on an expert's decision. Patients suffering from kidney disease, chronic liver disease, rheumatism, immunodeficiency disease, any type of cancer, inflammatory diseases such as inflammatory bowel disease, and infectious diseases in the past three months were excluded from the study. Patients with a medical history of any kind of surgery in the past three months, coronary angioplasty, and consumption of specific medications (such as steroids, penicillin, and oral contraceptives), those receiving hormone replacement therapy, and pregnant or lactating women were excluded, as well. Two checklists were used to evaluate the variables including information about the patients' medical records and laboratory results. Considering the first type error of 0.05, the second type error of 0.1, and odds ratio of 1.43 obtained for at least 964 angiography candidates and taking a 20% loss rate into account, the total sample size was estimated using the following formula.

N = 964 + 20% (964) = 1157

Finally, after 20 months, the data were collected from 1187 patients. After excluding the missing cases, 1128 patients were included in the study. It is worth mentioning that the variables with a large number of missing cases were not included in the model. The angiographic outcome (positive or negative) was considered the dependent variable. In addition, the most important variables in the diagnosis of CAD were gender, age, smoking habit, Body Mass Index (BMI), and Fasting Blood Sugar (FBS), which were selected based on a previous research and experts' opinions.

Generally, proper data scaling is very important to a dataset; otherwise, a variable may have a large impact on the prediction variable only because of its scale. Thus, before fitting the models in the present study, min-max normalization was employed to remove the scaling effects of all the variables.

3.2. Random Forest

RF refers to a combination of predictive trees, so that each tree depends on the values of an independent random vector sampled with the same distribution for all the trees in the forest. If the data set contains p number of potential predictor variables, $X = (X_1, X_2, \ldots, X_p)$ where $j = 1, 2, \ldots, p$ and Y is the feature under investigation (in this study, positive and negative angiographic results). Therefore, $X_j = (X_{1j}, X_{2j}, \ldots, X_{nj})$ and *n* is the sample size. If b is the number of trees and B is the total number of trees, the initial value of b is considered 1. In step k, θ_k number of independent samples with the same distribution from the data set are selected as the training set. $\underline{x}$ number of random samples are also selected from the predictive variables set. Then, the predictive function of $h(\underline{x}, \theta_k)\, h(\underline{x}, \theta_k)$ is created. Finally, the number of B trees is obtained by B repeating steps. If P > Q ($Q = \sum_{k=1}^{E} I(h(\underline{x}, \theta_k)) = 0$,$P = \sum_{k=1}^{E} I(h(\underline{x}, \theta_k)) = 1$) , the RF model will predict that $\underline{x}$ belongs to class 1 (positive result of angiography). If P < Q, the model will predict that $\underline{x}$ belongs to class 0 (negative result of angiography) (4).

The training and test sets are usually two-thirds and one-third of the sample size, respectively. In each split, the set size $\underline{x}$ (predictor variables) is randomly selected, which is equal to the square root of the number of variables ($\sqrt{P}$) (11). After fitting the model, the relative importance of each feature is measured using the Gini feature importance method. This method measures the number of nodes in the

tree that use that feature and reduces impurities throughout the forest trees. For each feature, scores are calculated after training and the results are compared, such a way that the sum of all importance values is equal to 1. Finally, the performance of the model is evaluated (12).

3.3. Artificial Neural Network

ANNs are computing systems inspired by biological neural networks. An ANN is based on a collection of connected units or nodes called artificial neurons and consists of three layers:

1. Input layer that takes inputs based on the existing data.
2. Hidden layer that uses backpropagation to optimize the weights of the input variables in order to improve the predictive power of the model.
3. Output layer that predicts the output based on the data from the input and hidden layers.

Input data are introduced to the neural network through the input layer that has one neuron for each component present in the input data and is communicated to hidden layers (one or more) in the network. These layers are called hidden, because they do not constitute the input or output layer. In the hidden layers, all the processing actually happens through a system of connections characterized by weights and biases. Once the input is received, the neuron calculates a weighted sum adding the bias. According to the result and an activation function (the most common one is sigmoid), it decides whether it should be 'fired' or 'activated.' Then, the neuron transmits the information downstream to other connected neurons in a process called 'forward pass.' At the end of this process, the last hidden layer is linked to the output layer, which has one neuron for each possible desired output.

Sigmoid neurons are modified perceptron. Like a perceptron, the sigmoid neuron has inputs x_1, x_2, …. However, instead of being just 0 or 1, these inputs can be any value between 0 and 1. A sigmoid neuron also has weights for each input, w_1, w_2, …, and an overall bias, b. Yet, the output is not 0 or 1. It is σ (w.x + b) where σ is called the sigmoid function:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

The output of a sigmoid neuron with inputs x_1, x_2, …, weights w_1, w_2, …, and bias b can be calculated using the following formula (13):

$$output = \frac{1}{1+\exp(-\sum_j w_j \cdot x_j - b)}$$

Here, similar to the RF model, the data are divided into a training set and a test set. The training set is used to find the relationships between the dependent and independent (important features) variables, while the test set analyzes the performance of the model.

3.4. Performance Evaluation Models

The performance of the RF and ANN models was evaluated based on the following criteria: recall, precision, accuracy F1 score, and Receiver Operating Characteristic (ROC) curve. At the end of the classification process, each case was placed in one of the following four groups:

- A True Positive (TP) is an outcome where the model correctly predicts the positive class.
- A False Negative (FN) is an outcome where the model incorrectly predicts the negative class.
- A True Negative (TN) is an outcome where the model correctly predicts the negative class.
- A False Positive (FP) is an outcome where the model incorrectly predicts the positive class.

A summary of the prediction results on a classification problem are presented by a confusion matrix ($M=\begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}$). This matrix is also a good way to check the performance of the classification models while new data are entered. If $TP+FN=n^+$ and $TN+FP=n^-$, a large number of true positives and negatives and a small number of false positives and negatives are expected. When a classification is performed completely in an ML algorithm, the confusion matrix is described as M ($M=\begin{pmatrix} n^+ & 0 \\ 0 & n^- \end{pmatrix}$).

Accuracy: The ratio of correctly predicted cases to the total cases in the data set.

$$Accuracy = \frac{TP+TN}{n^+ + n^-}$$

Sensitivity: The ratio of the number of correctly predicted positives (TPs) to the sum of the number of correctly predicted positives (TPs) and incorrectly predicted negatives (FNs).

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

Specificity: The ratio of the number of correctly predicted negatives (TNs) to the sum of the number of correctly predicted negatives (TNs) and incorrectly predicted positives (FPs).

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

Precision: The ratio of the number of correctly predicted positives (TPs) to the total number of predicted positives (true and false positives).

$$\Pr ecision = \frac{TP}{TP+FP}$$

ROC curve: A comprehensive assessment of the classification performance can be performed by the ROC curve. A ROC curve plots the classification results from the most positively to the most negatively classified items.

$$ROC = \frac{P(x|positive)}{P(x|negative)}$$

P (x|positive) represents the conditional probability that the input data have a positive class label (14). The Area Under the Rock Curve (AUC) is one of the most accurate model assessment criteria in classification problems, which indicates how well separated classes are based on the modeling algorithm. A value of 1 indicates a perfect fit and a value near 0.5 indicates that the model cannot

discriminate different groups. It is worth mentioning that prediction is done accidentally in this model (14).

$$AUC = (sensitivity + specificity) / 2$$

All analyses were performed using the R software (version 4.1.3), and the significance level was set at 0.05.

4. Results

Out of the 1128 patients, 752 (66.7%) presented positive angiography results. The distribution of the patients' demographic characteristics and other risk factors based on the angiography results have been presented in Table 1.

In this study, all the variables were scaled in the range of 0 to 1. The scaled data were used to fit the RF and ANN models. In these models, the training set was used to create the models and the test set was used to predict them. Two-thirds of the sample size were randomly selected as the training set (820 patients) and the rest as the test set (308 patients).

4.1. Random Forest Model's Performance

In the RF model, there were 100 trees and two variables were tried at each split. After fitting the model on the training set, classification error was obtained as 40.25%. According to the plot, the error rate was stabilized and decreased with the increase in the number of trees (Figure 1).

After fitting the model, the importance of the variables was explored. This is a fundamental outcome of RF and shows the importance of each variable in classifying the data. The mean decrease accuracy plot indicates the degree of accuracy that the model loses by excluding each variable. The more the accuracy suffers, the more important the variable is for successful classification. In the present study, FBS followed by gender, age group, smoking habit, and BMI were the most important predictor variables for an angiography candidate. The mean decrease in the Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting RF. The higher the value of mean decrease accuracy or mean decrease Gini score, the higher the importance of the variable in the model will be. In this study, FBS followed by gender, age group, BMI, and smoking habit were the most important predictive variables for an angiography candidate (Figure 2).

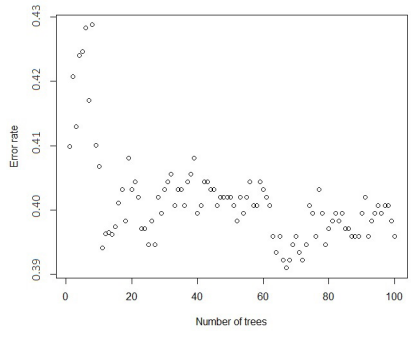

Figure 1. Error Rate of the Random Forest Model

The confusion matrix based on the test set demonstrated that 83 patients with positive angiography results were incorrectly predicted to have negative results (FNs) and 41 patients with negative angiography results were predicted to have positive results (FPs) (Table 2). Finally, the accuracy of the model was obtained as 59.74%.

4.2. Artificial Neural Network Model's Performance

In the ANN model, there were two repetitions for the neural network's training. In addition, there were three hidden neurons in each layer. The classification error was obtained as 27.20%. In this model, both repetitions were converged. Nonetheless, the output in the first repetition was used due to giving fewer errors (804.1761) compared to the second repetition (810.43). Overall, the ANN plot presented a model with five inputs, two outputs, and three nodes in a single hidden layer. Bias nodes connected to the hidden

Table 1. Comparison of the Frequency of the Patients' Demographics and Risk Factors Based on the Angiography Results

Variable	Angiography		P-value
	Positive	Negative	
Gender			
Female	283 (48.5)	262 (51.5)	(P < 0.001)
Male	460 (84.4)	123 (15.6)	
Age group (years)			(P < 0.001)
18 - 39	21 (37.5)	35 (62.5)	
40 - 59	374 (61.0)	239 (39.0)	
≥60	348 (75.8)	111 (24.2)	
Smoking habit			(P < 0.001)
Smoker	182 (75.5)	59 (24.5)	
Previous smoker	119 (69.5)	52 (30.5)	
Non-smoker	442 (61.7)	274 (38.3)	
Body mass index (kg/m^2)			(P = 0.003)
Underweight: < 18.5	12 (50)	12 (50)	
Normal weight: 18.5 - 25	249 (66.4)	126 (33.6)	
Overweight: 25 - 30	292 (65.6)	153 (34.4)	
Obesity: > 30	190 (66.9)	94 (33.1)	
Fasting blood sugar			(P < 0.001)
Normal: < 100	250 (60.6)	162 (39.4)	
Prediabetes: 100 - 125	163 (49.5)	160 (49.5)	
Diabetes: ≥ 126	330 (83.9)	63 (16.1)	

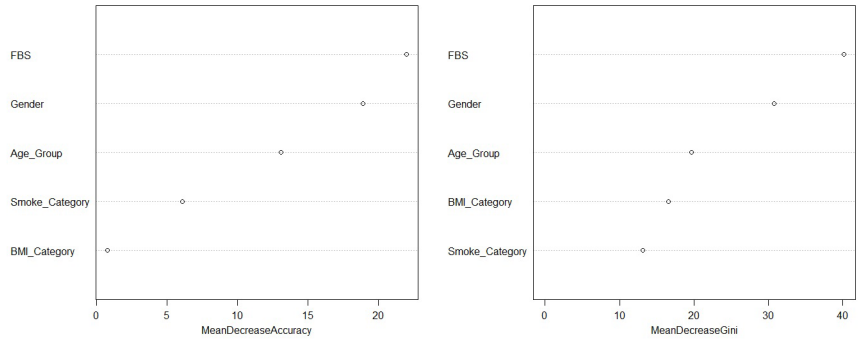

Figure 2. Importance of Features

Table 2. Confusion Matrix for Detecting the Angiography Result

Actual Class	Predicted Class			
	Artificial Neural Network		Random Forest	
	Negative	Positive	Negative	Positive
Negative	50	24	26	41
Positive	59	175	83	158

and output layers were also shown as number one. Lines connected to nodes indicated weighted connections between the layers. Besides, numbers on the lines represented the relative magnitude of each weight. Positive weights between the layers increased the net input, while negative weights lowered the net input (Figure 3).

The confusion matrix based on the test set demonstrated that 59 patients with positive angiography results were incorrectly predicted to have negative results (FNs) and 24 patients with negative angiography results were predicted to have positive results (FPs) (Table 2). Finally, the accuracy of the model was obtained as 73.05%.

4.3. The Models' Performance Evaluation Criteria

Sensitivity, specificity, and precision were reported for the RF and ANN models. The high-performance evaluation criteria of the ANN model revealed the good performance of this model compared to the RF model. Consequently, using the ANN model, individuals with positive angiography results were more likely to be correctly diagnosed for angiography in comparison to those with negative results (Table 3).

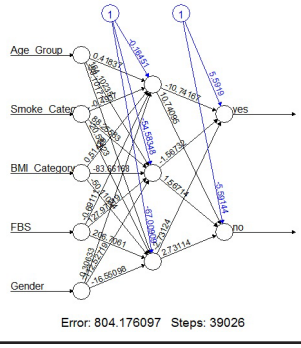

Figure 3. Artificial Neural Network Model's Performance

4.4. ROC and AUC Results

The performance of the models was determined by drawing an ROC curve and calculating the AUC. The AUC values for the test data set in the RF (0.52) and ANN (0.75) models indicated that the latter was much more accurate in classifying the dependent variable (Figure 4).

5. Discussion

This study aimed to compare the RF and ANN models to identify the most important features for predicting the need for angiography and positive CAD results. The variables inputted in the ML models were FBS, age, smoking habit, BMI, and gender. Among these five features, FBS was the most important factor predicting the necessity to perform angiography in the study population. FBS has been shown to be associated with the presence of CAD. In a study performed on 557 patients undergoing elective angiography, the results of multivariable analysis indicated that the odds of CAD was three times higher in patients with FBS >11.0 mmol/L compared to those with lower FBS levels. Therefore, FBS was mentioned as a strong independent

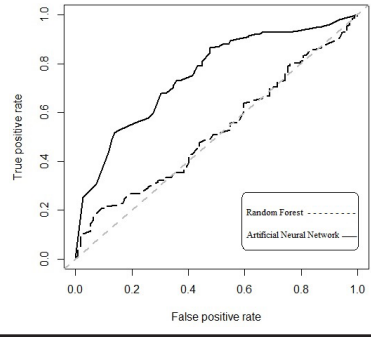

Figure 4. ROC Curve of the Random Forest and Artificial Neural Network Models

Table 3. Performance Evaluation Criteria for the Random Forest and Artificial Neural Network Models

	Sensitivity	Specificity	Precision
Artificial neural network	87.93	45.87	79.39
Random forest	79.39	23.85	65.56

predictor of CAD (15). Moreover, FBS was found to be associated with the severity of CAD. Using the Gensini score, Zhao et al. investigated the relationship between the severity of CAD and FBS in patients with confirmed CAD on angiography. They came to the conclusion that in non-diabetic patients, hyperglycemia and impaired fasting glucose were associated with a higher Gensini score and a more complex CAD (16). Similar results were also obtained in other studies (1). Consistently, the present study results revealed FBS as the most important feature in predicting CAD. Basic research works have demonstrated the effect of hyperglycemia on the development and progression of atherosclerosis and vascular damage through inducing reactive oxygen species production and endothelial dysfunction (17).

In the current research, BMI was found to be the second most important feature in predicting CAD. The previous studies have reported conflicting findings regarding the association between BMI and the presence and/or extent of CAD assessed by angiography or Coronary Computed Tomography Angiography (CCTA). Some studies showed that increase in BMI was linked with the presence of CAD (18-20), while some studies did not find any significant associations in this regard (15, 21). On the other hand, some studies demonstrated that BMI and obesity (BMI ≥ 30 kg/m^2) were inversely related to the risk of CAD (15, 22) and that patients with CAD were less likely to be obese (23). With regard to the association between BMI and the extent of CAD, the previous studies have rendered inconclusive results (20, 24, 25). Hence, many studies have suggested the utilization of other indices of central and visceral obesity (25-27). Yet, more prospective research is needed to identify the usefulness of BMI for CAD risk assessment in clinical practice.

Smoking habit, gender, and age were the other important factors that had a predictive utility for the presence of CAD. The present study results were in agreement with those of the previous studies, which indicated that male patients, old ones, and current or previous smokers were at a greater risk of having CAD on angiography (28, 29). These risk factors were also used to identify the patients with high-risk CAD defined as severe left main or proximal left anterior descending artery stenosis (30). Evidence has also highlighted the predictive role of smoking in CAD diagnosis (31, 32). In addition, the duration of smoking was an important factor in predicting the presence of CAD in both current and previous smokers (31). Moreover, the number of smoked cigarettes was found to be significantly higher in patients with CAD (33, 34). These findings concur with the fact that smoking-induced inflammation can lead to both atherosclerotic plaque development and progression. Besides, smokers are at a higher risk of coronary events due to inflammation and hypercoagulability state (35).

Despite the higher prevalence of cardiovascular risk factors among females, the incidence of CAD was higher in males. Male patients were twice more likely to have positive angiography results for CAD compared to females (31). Generally, females have less extensive yet high risk plaques. They are at a greater risk of plaque erosion and atherosclerosis progression after menopause, since estrogen plays a critical role in halting plaque development and aiding plaque stabilization (36). On the other hand, the risk of CVDs is increased with advanced age due to changes in blood vessels and arterial stiffness (31). A prior study disclosed that patients older than 65 years were at a 15-fold higher risk of CAD compared to those aged below 45 years (29).

The decision on whether a patient needs invasive angiography is complex and many risk factors must be considered. ML approaches can construct models to screen candidates for angiography and accelerate the process of medical decision-making. Sang-Yeong Cho et al. carried out a study on Korean adults regarding the development of ML-based risk prediction models (logistic regression, tree bag, RF, and the adaboost neural network model) and compared their performances with the pre-existing algorithms (Framingham Risk Score (FRS), Pooled Cohort Equation (PCE), Systematic Coronary Risk Evaluation (SCORE), and QRISK3). The results showed that the models using the ML techniques improved cardiovascular risk prediction. Among the ML algorithms, neural networks enabled the learning of highly complex functions and accurate predictions for complex decision-making problems (37). Yuqing Tian et al. also conducted a study to construct a genetic diagnostic model of heart failure using RF and ANN. The RF algorithm was used to identify the key genes expressed in heart failure. Then, these key genes were inputted in ANN in order to build a genetic diagnostic model of heart failure. The results revealed the excellent AUC efficiency of ANN (38). In another research, Joloudari et al. compared four ML methods, namely random trees, decision tree of C5.0, support vector machine, and decision tree of Chi-squared Automatic Interaction Detection (CHAID), for CAD diagnosis. Based on the results, the random trees model showed a superior performance compared with others (39). Stephen F Weng et al. also compared four ML methods (logistic regression, RF, gradient boosting machines, and neural networks) to determine the potential of ML in improving cardiovascular risk prediction by using routine clinical data. The performance of the models was investigated by AUC, sensitivity, specificity, positive predictive value, and negative predictive value. Among these four algorithms, neural network indicated the best performance, which was in line with the current study findings (40). In another investigation, Muhammad Saqib Nawaz et al. applied six ML algorithms (gradient descent optimization, k-nearest neighbor, naïve Bayes, ANN, RF, and support machine vector) for intelligent cardiovascular disease prediction. According to the results, the best performance was shown by gradient descent optimization, with the accuracy, sensitivity, and precision of 98.54%, 99.43%, and 97.76%, respectively. Additionally, comparison of RF and ANN revealed the latter's better performance, which was consistent with the findings of the current research (41).

5.1. Conclusion

The study results demonstrated that FBS, gender, age, BMI, and smoking habit were important features in predicting the results of an angiography for CAD. Thus,

applying these factors in ML approaches can be considered a screen for angiography to accelerate the treatment process. Furthermore, the ANN algorithm is a powerful method for prediction, as it mimics the biological neural networks. However, multi-center studies using hybrid ML methods are recommended to make a stronger predictive model.

5.2. Ethical Approval

The study design and protocols were approved by the Ethics Committee of the Research Vice-chancellor of Mashhad University of Medical Sciences (code: IR.MUMS.REC.1399.357).

5.3.Informed Consent

Written informed consent forms were obtained from the patients.

Acknowledgements

The authors would like to thank the Research Vice-chancellor of Mashhad University of Medical Sciences, Mashhad, Iran for financially supporting this project (research project code: 981532).

Authors' Contribution

Conceptualization: S.ST. and P.G.; methodology: S.ST.; software: S.ST. and P.G.; validation: P.G., M.T., and M.M.; formal analysis: P.G., S.ST., and M.T.; investigation: M.GM., A.T., N.A., and P.G.; resources: M.GM., M.T., A.T., and M.M.; data curation: P.G.; original draft preparation: P.G. and S.ST.; draft review and editing: S.ST. and M.B.; visualization: S.ST.; supervision: S.ST. All authors have read and confirmed the published version of the manuscript.

Funding/Support

This study was supported by grant No. 981532 from the Research Vice-chancellor of Mashhad University of Medical Sciences.

Financial Disclosure

The authors have no financial interests related to the material in the manuscript.

References

1. Qin Y, Yan G, Qiao Y, Ma C, Liu J, Tang C. Relationship between Random Blood Glucose, Fasting Blood Glucose, and Gensini Score in Patients with Acute Myocardial Infarction. BioMed research international. 2019;2019:9707513.
2. Mahmoodabadi Z, Abadeh MS. CADICA: Diagnosis of coronary artery disease using the imperialist competitive algorithm. Journal of Computing Science and Engineering. 2014;8(2):87-93.
3. Noh D-W, Kim S. Associations between coronary artery stenosis detected by coronary computed tomography angiography and the characteristics of health checkup examinees in the Republic of Korea. Radiography. 2020;26(1):22-6.
4. Breiman L. Random Forest. Machine Learning2001.
5. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: Springer; 2013.
6. Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. International Journal of Computer Science Issues (IJCSI). 2012;9(5):272.
7. Tayefi M, Tajfard M, Saffar S, Hanachi P, Amirabadizadeh AR, Esmaeily H, et al. hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. Computer methods and programs in biomedicine. 2017;141:105-9.
8. Elham Shamsara SSS, Mohammad Tajfard, Ivan Yamshchikov, Habibollah Esmaily*, Maryam Saberi-Karimian, Hamideh Ghazizadeh, Seyed Reza Mirhafez, Zahra Farjami, Gordon A. Ferns and Majid Ghayour-Mobarhan*. Artificial Neural Network Models for Coronary Artery Disease, . 2021;16 (4):610 - 23.
9. Ohri A. Python® for R users : a data science approach. the United States of America: John Wiley & Sons; 2018.
10. Renganathan V. Overview of artificial neural network models in the biomedical domain. Bratislavske lekarske listy. 2019;120(7):536-40
11. Gareth J, Daniela W, Trevor H, Robert T. An introduction to statistical learning: with applications in R: Spinger; 2013.
12. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC bioinformatics. 2009;10(1):213.
13. Haykin S. Neural networks and learning machines, 3/E: Pearson Education India; 2009.
14. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition. 1997;30(7):1145-59.
15. Al-Shudifat AE, Azab M, Johannessen A, Al-Shdaifat A, Agraib LM, Tayyem RF. Elevated fasting blood glucose, but not obesity, is associated with coronary artery disease in patients undergoing elective coronary angiography in a referral hospital in Jordan. Annals of Saudi medicine. 2018;38(2):111-7.
16. Zhao T, Gong HP, Dong ZQ, Du YM, Lu QH, Chen HQ. Predictive value of fasting blood glucose for serious coronary atherosclerosis in non-diabetic patients. The Journal of international medical research. 2019;47(1):152-8.
17. Shah MS, Brownlee M. Molecular and cellular mechanisms of cardiovascular disorders in diabetes. Circulation research. 2016;118(11):1808-29.
18. Dores H, de Araújo Gonçalves P, Carvalho MS, Sousa PJ, Ferreira A, Cardim N, et al. Body mass index as a predictor of the presence but not the severity of coronary artery disease evaluated by cardiac computed tomography. European journal of preventive cardiology. 2014;21(11):1387-93.
19. Nafakhi H, Al-Mosawi AA, Al-Buthabhak K. Sex-Related Differences in the Association of BMI and Pericardial Fat Volume With Coronary Atherosclerotic Markers in Young. Angiology. 2021;72(3):285-9.
20. Labounty TM, Gomez MJ, Achenbach S, Al-Mallah M, Berman DS, Budoff MJ, et al. Body mass index and the prevalence, severity, and risk of coronary artery disease: an international multicentre study of 13 874 patients. European Heart Journal–Cardiovascular Imaging. 2013;14(5):456-63.
21. Cho JH, Kim H-L, Kim M-A, Oh S, Kim M, Park SM, et al. Association between obesity type and obstructive coronary artery disease in stable symptomatic postmenopausal women: data from the KoRean wOmen'S chest pain rEgistry (KoROSE). Menopause (New York, NY). 2019;26(11):1272-6.
22. Bechlioulis A, Vakalis K, Naka KK, Bourantas CV, Papamichael ND, Kotsia A, et al. Paradoxical protective effect of central obesity in patients with suspected stable coronary artery disease. Obesity. 2013;21(3):E314-E21.
23. Phillips SD, Roberts WC. Comparison of body mass index among patients with versus without angiographic coronary artery disease. The American journal of cardiology. 2007;100(1):18-22.
24. Parsa AFZ, Jahanshahi B. Is the relationship of body mass index to severity of coronary artery disease different from that of waist-to-hip ratio and severity of coronary artery disease? Paradoxical findings: cardiovascular topic. Cardiovascular Journal of Africa. 2015;26(1):13-6.
25. Lewandowski A, Dłużniewski M, Chmielewski M, Zieliński Ł, Pikto-Pietkiewicz W, Burbicka E, et al. Evaluation of the relations between the presence of the metabolic syndrome and the degree of visceral obesity and the severity of coronary artery disease by coronary angiography. Kardiologia Polska (Polish Heart Journal). 2013;71(9):937-44.
26. Sabah KMN, Chowdhury AW, Khan HLR, Hasan AH, Haque S, Ali S, et al. Body mass index and waist/height ratio for prediction of severity of coronary artery disease. BMC research notes. 2014;7(1):1-7.

27. Rubinshtein R, Halon DA, Jaffe R, Shahla J, Lewis BS. Relation between obesity and severity of coronary artery disease in patients undergoing coronary angiography. The American journal of cardiology. 2006;97(9):1277-80.
28. Azab M, Al-Shudifat A-E, Johannessen A, Al-Shdaifat A, Agraib LM, Tayyem RF. Are Risk Factors for Coronary Artery Disease Different in Persons With and Without Obesity? Metabolic syndrome and related disorders. 2018;16(8):440-5.
29. Al-Shudifat A-E, Johannessen A, Azab M, Al-Shdaifat A, AbuMweis SS, Agraib LM, et al. Risk factors for coronary artery disease in patients undergoing elective coronary angiography in Jordan. BMC Cardiovasc Disord. 2017;17(1):183.
30. Jang JJ, Bhapkar M, Coles A, Vemulapalli S, Fordyce CB, Lee KL, et al. Predictive Model for High-Risk Coronary Artery Disease. Circulation Cardiovascular imaging. 2019;12(2):e007940.
31. Noh DW, Kim S. Associations between coronary artery stenosis detected by coronary computed tomography angiography and the characteristics of health checkup examinees in the Republic of Korea. Radiography (London, England : 1995). 2020;26(1):22-6.
32. Buljubasic N, Akkerhuis KM, de Boer SP, Cheng JM, Garcia-Garcia HM, Lenzen MJ, et al. Smoking in Relation to Coronary Atherosclerotic Plaque Burden, Volume and Composition on Intravascular Ultrasound. PloS one. 2015;10(10): e0141093.
33. Yano M, Miura S, Shiga Y, Miyase Y, Suematsu Y, Norimatsu K, et al. Association between smoking habits and severity of coronary stenosis as assessed by coronary computed tomography angiography. Heart and vessels. 2016;31(7):1061-8.
34. Kim JA, Chun EJ, Lee MS, Kim KJ, Choi SI. Relationship between amount of cigarette smoking and coronary atherosclerosis on coronary CTA in asymptomatic individuals. The international journal of cardiovascular imaging. 2013;29 Suppl 1:21-8.
35. Yasue H, Hirai N, Mizuno Y, Harada E, Itoh T, Yoshimura M, et al. Low-grade inflammation, thrombogenicity, and atherogenic lipid profile in cigarette smokers. Circulation Journal. 2006;70(1):8-13.
36. Lansky Alexandra J, Ng Vivian G, Maehara A, Weisz G, Lerman A, Mintz Gary S, et al. Gender and the Extent of Coronary Atherosclerosis, Plaque Composition, and Clinical Outcomes in Acute Coronary Syndromes. JACC: Cardiovascular Imaging. 2012;5(3_Supplement):S62-S72.
37. Cho S-Y, Kim S-H, Kang S-H, Lee KJ, Choi D, Kang S, et al. Pre-existing and machine learning-based models for cardiovascular risk prediction. Scientific reports. 2021;11(1):1-10.
38. Tian Y, Yang J, Lan M, Zou T. Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure. Aging (Albany NY). 2020;12(24):26221.
39. Joloudari JH, Hassannataj Joloudari E, Saadatfar H, Ghasemigol M, Razavi SM, Mosavi A, et al. Coronary artery disease diagnosis; ranking the significant features using a random trees model. International journal of environmental research and public health. 2020;17(3):731.
40. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PloS one. 2017;12(4):e0174944.
41. Nawaz MS, Shoaib B, Ashraf MA. Intelligent Cardiovascular Disease Prediction Empowered with Gradient Descent Optimization. Heliyon. 2021;7(5):e06948.