Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2023 (May);37.46. https://doi.org/10.47176/mjiri.37.46

Assessment of the Performances of Adaptive Neuro-Fuzzy Inference System and Two Statistical Methods for Diagnosing Coronary Artery Disease

Zahra Rajabzadeh[1], Nooshin Akbari Sharak[2], Habibollah Esmaeili[3], Majid Ghayour-Mobarhan[4,5], Mohammad Taghi Shakeri[6]*, Zahra Pasdar[7]

Received: 22 Jan 2022 Published: 2 May 2023

Abstract

Background: The accurate diagnosis of cardiac disease is vital in managing patients' health. Data mining and machine learning techniques play an important role in the diagnosis of heart disease. We aimed to examine the diagnostic performances of an adaptive neuro-fuzzy inference system (ANFIS) for predicting coronary artery disease and compare this with two statistical methods: flexible discriminant analysis (FDA) and logistic regression (LR).

Methods: The data of this study is the result of descriptive-analytical research from the study of Mashhad. We used ANFIS, LR, and FDA to predict coronary artery disease. A total of 7385 subjects were recruited as part of the Mashhad Stroke and Heart Atherosclerotic Disorders (MASHAD) cohort study. The data set contained demographic, serum biochemical parameters, anthropometric, and many other variables. To evaluate the ability of trained ANFIS, LR, and FDA models to diagnose coronary artery disease, we used the Hold-Out method.For analyzing data, we used SPSS v25, R 4.0.4, and MATLAB 2018 software.

Results: The accuracy, sensitivity, specificity, Mean squared error (MSE) , and area under the roc curve (AUC) for ANFIS were 83.4%, 80%, 86%, 0.166 and 83.4%. The corresponding values based on the LR method were 72.4%, 74%, 70% , 0.175 and 81.5% and for the FDA method, these measurements were 77.7%, 74%, 81%, 0.223, and 77.6%, respectively.

Conclusion: There was a significant difference between the accuracy of these three methods. The present findings showed that ANFIS was the most accurate method for diagnosing coronary artery disease compared with LR and FDA methods. Thus, it could be a helpful tool to aid medical decision-making for the diagnosis of coronary artery disease.

Keywords: Adaptive Neuro-Fuzzy Inference System, Logistic Regression, Flexible Discriminant Analysis, Coronary Artery Disease

Conflicts of Interest: *None declared*
Funding: *Mashhad University of Medical Science*

Cite this article as: Rajabzadeh Z, Akbari Sharak N, Esmaeili H, Ghayour-Mobarhan M, Shakeri MT, Pasdar Z. Assessment of the Performances of Adaptive Neuro-Fuzzy Inference System and Two Statistical Methods for Diagnosing Coronary Artery Disease. *Med J Islam Repub Iran.* 2023 (2 May);37:46. https://doi.org/10.47176/mjiri.37.46

Introduction

Coronary Artery Disease (CAD) is the most common type of heart disease, and it is caused by the accumulation of plaque, namely cholesterol deposits among other substances, in the walls of the heart's blood vessels (coronary

Corresponding author: Dr Mohammad Taghi Shakeri, ShakeriMT@mums.ac.ir

1. Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran
2. Student Research Committee, Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran
3. Rheumatic Diseases Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
4. Metabolic Syndrome Research Center, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
5. International UNESCO Center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, Iran
6. Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
7. Institute of Applied Health Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen AB25 2ZD, UK

↑What is "already known" in this topic:
Coronary Artery Disease is the most common type of heart disease and a major worldwide health problem with overwhelmingly high incidence and mortality rates. Diagnosing heart disease is a complex and time-consuming task in medical science, and timely diagnosis can decrease mortality hazards. So Diagnostic techniques play an important role in health and medical research.

→What this article adds:
In this article, we implemented an adaptive neuro-fuzzy inference system for predicting coronary artery disease and compared it with flexible discriminant analysis and logistic regression.

arteries). Over time, this narrows the coronary arteries and can partially or completely block blood flow. This can lead to angina and heart attack due to lack of perfusion and consequently, oxygenation, which is required to meet its demands. In the 21st century, CAD is still the leading cause of death in developed countries and is, therefore, the leading cause of death worldwide (1). CAD is a major worldwide health problem with overwhelmingly high incidence and mortality rates (2). According to the World Health Organization (WHO), 17.9 million die as a result of the cardiovascular disease annually, accounting for 31% of the world's death (3). The Institute of Standards and Health Assessment at the University of Washington reported a 27.4 percent increase in deaths from the disease in Iran between 2007 to 2017 (4).

In Iran, cardiovascular diseases account for 50% of annual mortality (1), which includes ischemic heart disease (26% in men and women), stroke (10% in men and 13% in women), heart disease with high blood pressure (5% in men and women) and other heart diseases (5% in men and 6% in women) (2).

Diagnosing heart disease is a complex and time-consuming task in medical science. If a heart disease is not diagnosed in time, the lives of people with it are endangered and the mortality rate increases. Therefore, the timely diagnosis of heart disease is a pressing matter. Current diagnostic tests for heart disease include heart tests, blood tests and ECGs which are time-consuming. Most of the time, diagnoses and decisions are made by doctors, but experienced doctors are not available everywhere and at all times, so we need a system that can accurately and quickly predict heart disease (5). Early detection of cardiovascular disease helps physicians determine the most appropriate treatment and increase patients' chances of survival (6).

Diagnostic techniques play a very important role in health and medical research.

For numerous practical CAD classification tasks, traditional parametric statistical models are often used (4, 7). One of the main purposes of modeling and classifying statistics is to predict based on available facts, variables, and available information on a particular topic. In statistics, this task was performed by methods such as regression, discriminant analysis, regression tree, and other statistical methods.parametric statistical models for modeling the relationships between variables have several assumptions and limitations. These include issues with missing values, sample size and distribution restrictions. In practice, if the actual data does not meet the assumptions of the model, the use of these methods is impossible or may provide misleading results with significant errors. In addition, none of these methods are capable of modeling complex nonlinear relationships and high-degree interactions. Therefore, there is a need for methods that are less restrictive in this area. In medical sciences, methods based on data mining, artificial intelligence methods, fuzzy method and so forth are used for accurate and timely diagnosis of heart disease (5, 7). Artificial neural networks (ANNs), quadratic discriminant analysis (QDA) and logistic regression (LR) and ANFIS are methods available in the medical field and have been used in the diagnosis of several diseases as well as CAD. In the study of Ziasabonchi and his colleagues, the aim was to create a method for classifying heart patients based on the characteristics of the patients using the adaptive neural fuzzy inference system method. The results showed that the accuracy of the prediction model was 92.3% (8), Zabah et al. conducted a study to diagnose coronary artery disease using the fuzzy neuro method (9), Kurt et al. conducted research entitled "Comparison of the performance of logistic regression, classification tree, tree regression and neural network on the prediction of coronary artery disease". After comparing the performance of the models, the neural network with the area under the curve of 0.78% was recognized as the best model (10). Presi et al. used the fuzzy optimization technique to predict coronary heart disease using a fuzzy-based system (11), Mohammadpour et al used the artificial neural network model to predict coronary artery disease (12), Polat et al used principal component analysis and adaptive neural-fuzzy inference system (ANFIS) to diagnose diabetes, which is a very common and important disease(13). Mahmoudi et al.'s propose a genetic selection model based on classification model with fuzzy neural method (ANFIS) for cancer classification (14), Xu et al.'s used the logistic regression model to diagnose CAD (15). However, according to the different methods, their predictive accuracies might differ. In this paper, we compared the predictive accuracies of ANFIS trained with a hybrid learning algorithm (Least Squares Estimator and Back-Error Propagation) and Logistic regression (LR) and Flexible Discriminant Analysis (FDA) in predicting CAD.

Methods

Study population

The data of this study is the result of descriptive-analytical research from the study of Mashhad. The Mashhad stroke and atherosclerotic heart disorder (MASHAD) study was initiated at Mashhad University of Medical Sciences (MUMS), and a total of 9704 subjects were enrolled between 2001 and 2020. Variables measured in this database included demographic characteristics (including gender, age, level of education, family history, etc.), anthropometric information (including weight, height, body mass index, etc.), and biochemical parameters (triglyceride, total cholesterol, fasting blood sugar, systolic and diastolic blood pressure, etc.). The anxiety and depression scores were obtained from the Beck Anxiety Inventory and the Beck Depression Inventory, respectively. The determination of risk factors in the present study was done by reviewing articles related to coronary artery disease and consulting with experts. At this stage, most of the existing variables related to the disease were selected and analyzed in terms of having the necessary quality for analysis. In other words, the data were checked and corrected from aspects such as the completeness or absence of missing observations. In this study, people were divided into two groups according to the target variable (people with coronary artery disease and people without coronary artery disease).

Data from the MASHAD cohort were randomly partitioned into training (70%) and test (30%) data sets. The diagnostic function of the adaptive neural fuzzy inference system (ANFIS) to predict coronary artery disease was

[DOI: 10.47176/mjiri.37.46]

compared with two statistical methods: flexible discriminant analysis (FDA) and logistic regression (LR) analysis.

The 21 variables, including quantitative and qualitative variables, were included in the SPSS database. 20 variables were selected as input or predictor variables and one variable as the final or target variable.

Imbalanced datasets were a challenging problem in classification as they reduced the classification performance. Therefore, we used the Synthetic Minority Oversampling TEchnique (SMOTE) method to balance distribution in the class variable, which was defined as whether the subject is diagnosed with the disease or not. The SMOTE algorithm generates new samples in the neighborhood of existing samples using the Interpolation method (Equation 1) sampling approach based on the specificity space similarities between the minority samples.

$$x_{new} = x_i + (\hat{x_l} - x_i) * \delta$$

For each x_i in the minority class, the k nearest neighbor based on the Euclidean distance is identified. Then, to create a new artificial instance (x_{new}), one of the k neighbors next to the x_i is randomly selected ($\hat{x_l}$). According to equation 1, the vector calculates the difference between the corresponding properties of these two samples and multiplies it by a random number δ between [0, 1].

Finally, it adds the result to the x_i instance. This generates a new (x_{new}), or synthetic instance as part of the connecting line between the x_i and x_l(16).

Sensitivity and specificity were two indicators for evaluating the result of a classification experiment. Sensitivity means the proportion of sick people whose tests have correctly diagnosed the patient. Therefore, the larger value, the more accurate the diagnosis of the patient. Specificity means the proportion of healthy people that the model has correctly identified as healthy. The measure of accuracy was the ratio of the number of healthy and sick people who were correctly diagnosed. ROC curves are widely used in medical decision-making for diagnostic tests. The Area under the ROC Curve (AUC) is an indicator for measuring the performance of the method used in classification.

The theory of ANFIS, FDA, and LR models have been stated briefly, respectively. For analyzing data, we used SPSS v25, R 4.0.4, and MATLAB 2018 software.

ANFIS Model

The Adaptive Neural Fuzzy Inference System (ANFIS) is a new intelligent algorithm that combines both neural networks and fuzzy logic principles(17). It computed initial membership functions by using training data and afterward used a hybrid-learning algorithm to adjust membership functions. Its inference was based on fuzzy IF-THEN rules (18). Figure 1 shows the ANFIS architecture. The inputs were represented as X and Y, and each of them had two membership values. Also, the output was represented as f. Circles and squares were used to represent fixed nodes and the nodes that have parameters to learn, respectively. Suppose that the rule base contains two fuzzy if-then rules of Takagi and Sugenos type (17, 18):

Rule 1: If x is A_1 and y is B_1 then $f_1 = \alpha_1$ x+ β_1y+ γ_1

Rule 2: If x is A_2 and y is B_2 then $f_2 = \alpha_2$ x+ β_2y+ γ_2

where x and y are inputs, A_1, A_2 , B_1 and B_2 are fuzzy sets (which are determined during the learning process of the network), f_1 and f_2 are outputs α_2, α_1 ,β_1 , β_2 , γ_1 and γ_2 are linear parameters (which are determined during the learning process of the network) Are.

The ANFIS system has 5 layers which we expressed as follows:

Layer 1: Named the Fuzzy Layer, took the input values and determined the membership functions as follows:

$$O_i^1 = \mu_{Ai}(x). \quad i = 1,2. \qquad O_i^2 = \mu_{Bi}(x). \quad i = 1,2$$

Membership functions could be Gaussian or bell-shaped. The bell-shaped membership functions are suitable for nonlinear systems and are defined as:

$$\mu_{Ai}(x) = \frac{1}{1+|\frac{x-c_i}{a_i}|^{2b_i}} \qquad \text{and} \qquad \mu_{Bi}(y) = \frac{1}{1+|\frac{x-c_i}{a_i}|^{2b_i}}$$

where x and y are the input variables and three variable parameters (a_i, b_i و c_i) are related to the bell membership function, and they are called Premise parameters. And respectively, a is the width, b is the slope, and c is the center of the bell function, and these parameters can be determined and adjusted by the learning algorithm in the training process.Layer 2: Named as the Product Layer, consists of fixed nodes that perform as a simple multiplier and multiply incoming signals and sends the product out, generating the firing strengths for the rules. Its output is represented as:

$$O_i^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(y) = O_i^1 * O_i^2 \qquad i = 1,2$$

Layer 3. Named as the Normalized Layer. This has fixed nodes and is responsible for normalizing the firing strengths that have been computed in the previous layer.

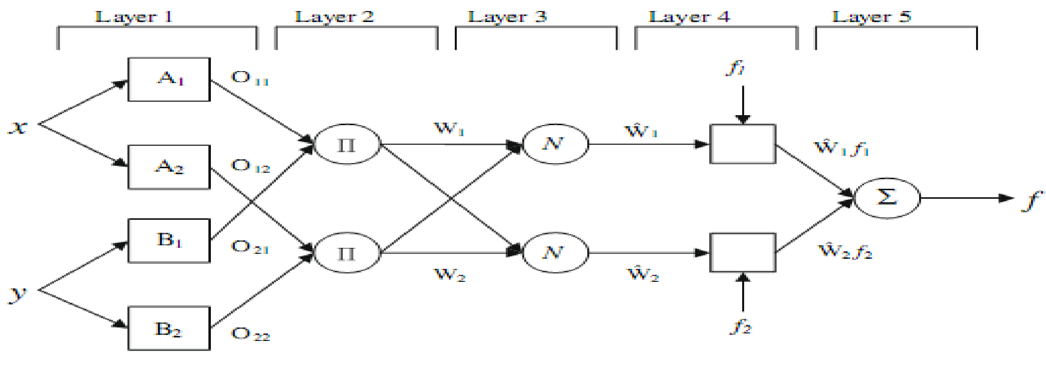

Figure 1. ANFIS architecture

The output is represented as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{\Sigma_i w_i}. \qquad i = 1,2$$

Layer 4. Named Defuzzy, it multiplies the normalized values from the previous layer and a first-order polynomial (p_i ,q_i ,r_i), and is represented as:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i). \qquad i = 1,2$$

Where $\bar{w}_i$ is the normalized weight from the previous layer and p_i, q_i and $r_i$are parameters (Consequent or result).

Layer 5. The Final Output Layer computes the overall output as the summation of all incoming signals, i.e.

$$O_i^5 = f = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\Sigma_i w_i}. \qquad i = 1,2$$

ANFIS has two adaptive layers, the first and fourth (17-19).

In the ANFIS structure, the final output can be expressed as a linear combination of the consequent parameters $\{p_1 . q_1 . r_1 . p_2 . q_2 \text{ and } r_2\}$ as follows:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2$$
$$= \bar{w}_1 f_1 + \bar{w}_2 f_2$$
$$= \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2)$$
$$= (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + \bar{w}_1 r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + \bar{w}_2 r_2$$

In the forward pass of the learning algorithm, consequent parameters are identified by the least squares estimate. In the backward pass, the error signals, which are the derivatives of the squared error with respect to each node output, propagate backward from the output layer to the input layer. In this backward pass, the premise parameters are updated by the gradient descent algorithm (20, 21). Two passes in the hybrid learning algorithm for ANFIS have been demonstrated in Table 1.

Logistic Regression (LR) is usually used when the independent variables include both numerical and nominal measures and the outcome variable is binary. LR can also be used when the outcome has more than two values. The LR model is expressed as:

$$p = P_r(y_r = 1) = \frac{e^{\alpha + \beta_1 x_{1.i} + \beta_2 x_{2.i} + \cdots \beta_k x_{k.i}}}{1 + e^{\alpha + \beta_1 x_{1.i} + \beta_2 x_{2.i} + \cdots \beta_k x_{k.i}}}$$

where $x_{1.i}, x_{2.i}. . ., x_{k.i}$ are independent variables, α the intercept, $\beta_1, \beta_2, . ., \beta_k$ the regression coefficients and exp indicate the base of the natural logarithm (e=2.718), which is taken to the power shown in square brackets. In logistic regression, the χ^2 test is used (instead of the t or F test) to test the fit of the model and the significance of the variables to determine whether a variable added significantly contributes to the prediction. In logistic regression, the odds ratio (p / (1-p)) is used, which is the ratio of the probability of success to the probability of failure, and the logarithm of the odds ratio is calculated based on the following equation. This model is known as the Logit model (22):

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_{1.i} + \beta_2 x_{2.i} + \cdots \beta_k x_{k.i} \quad i = 1.2. \dots . n$$

Flexible Discriminant Analysis (FDA) is defined based on the Linear discriminant analysis (LDA) linear audit analysis and developed by Fisher in 1936 (23). The LDA selects several variables from a large number of variables and identifies and distinguishes the groups in the response variable. Now, if nonlinear equations are predominant in a relationship, a series of flexible equations are used instead of linear equations. In this method, not only flexible linear equations but also nonlinear equations, such as quadratic, tertiary and logarithmic, are used and studied. Nonlinear prediction combinations such as splines were used for classification and finally, the best equation was selected. FDA replaced the linear regression step with a non-parametric or semi-parametric regression (24). One approach is to use Multivariate Adaptive Regression Splines (MARS) to fit the model. The quality of MARS is similar to that of tree-based models.MARS performed a forward stage of splitting and then, like tree models, pruned the model terms. MARS chose one or more predictors to "split" on. Multiple splits were used to model different predictors, but if a predictor was never used in a split, the class boundaries were functionally independent of that predictor. Thus, FDA simultaneously built a classification model while conducting feature selection (24).

Results

Data Description and Statistical Methods

The data of this study is the result of descriptive-analytical research from the study of Mashhad. Finally, 7385 complete observations were used for analysis, 235 of whom had coronary artery disease.

The distribution of features was not normal by the Kolmogorov–Smirnov (Lilliefors significant correction) test. The initial relationship between the target and 20 variables was performed by the Mann-Whitney U test and the chi-square test (χ^2) as appropriate. Variables with a p-value less than 0.2 were selected as input variables to the models. Finally, 18 variables were selected as inputs to the machine learning algorithms and LR .

Therefore, to perform the analysis and to ensure the results of the tests relate to the statistical hypotheses of the research question, data simulation was performed in the patient group using the SMOTE method. In this way, the two groups were adjusted in terms of sample size, and it was possible to test the hypotheses and fit the statistical models defined in this study. After simulation, the total number of people in the study reached 13025 people. From this data, 5875 people were in the group of coronary artery patients, CAD (+), and 7150 people were in the group of no CAD, CAD (-) patients.

Table 1. Two passes in the hybrid learning procedure for ANFIS

Parameters	Forward Pass	Backward Pass
Premise parameters	Fixed	Gradient descent
Consequent parameters	Least square	Fixed
Signals	Node outputs	Error signals

[DOI: 10.47176/mjiri.37.46]

Data Analysis

Descriptive statistics for quantitative and qualitative demographic, clinical, and biochemical characteristics of the study population have been reported in Tables 2 and 3, respectively. The mean age was 54.3±6.86 years for the CAD (+) group and 48.29±4.67 years for the CAD (-) group. The statistical analysis for indication of an initial association between CAD and 20 independent variables was conducted. These results indicated a significant association between CAD and age, body mass index (BMI), waist circumferences (WC), depression, anxiety, systolic blood pressure (SBP), Diastolic blood pressure (DBP), Triglyceride (TG), Fasting blood glucose (FBG), High-density lipoprotein (HDL), Low-density lipoprotein (LDL), High-sensitivity c-reactive protein (HS-CRP), Total Cholesterol (TC), physical activity level (PAL), family history, educational attainment, job and smoking at the significant level of 0.2. These variables formed the candidate variables and were entered into the models. The diagnostic function of the adaptive neural fuzzy inference system (ANFIS) to predict coronary artery disease was compared with two statistical methods: flexible discriminant analysis (FDA) and logistic regression (LR). A comparison of the Sensitivity, Specificity, Mean Square Error (MSE), and Accuracy rates for classification techniques has been demonstrated in Table 4. The confidence interval of sensitivity and specificity for classification techniques have been demonstrated in Table 5.

As shown in Table 4, the MSE value of the ANFIS with hybrid algorithm had a smaller constant than the LR and FDA/MARS algorithms. CAD was classified with accuracy rates varying from 72% to 83.4% by the ANFIS, LR, and FDA/MARS algorithms. Among the three different algorithms, ANFIS achieved the highest accuracy rate (83.4%). The accuracy rates of the statistical methods LR (72.4%) and FDA/MARS (77.7%) were poorer than the ANFIS. AUCs were 83.4% for ANFIS, 81.5% for LR and 77.7% for FDA/MARS. ROC curves of the ANFIS and statistical methods are shown in Figure 2. The AUC of ANFIS was higher than the AUCs of the LR and FDA. Comparisons on the accuracy of ANFIS, LR, and DA models are

Table 2. Descriptive statistics for quantitative clinical and biochemical characteristics of the study population

Variable	Mean ± SD (CAD +)	Mean ± SD (CAD -)	*P*-value
Serum biochemical parameters			
TG (mg/dl)	172.6 ± 104.65	141.46 ± 88.50	(<0.001)**
TC (mg/dl)	199.85 ± 41.56	190.97 ± 37.89	(<0.001) (0.001)**
LDL-C (mg/dl)	122.29 ± 35.50	116.51 ± 34.69	(0.012)**
HDL-C (mg/dl)	40.06 ± 9.07	42.73 ± 9.86	(0.001)**
Hs-CRP (mg/dl)	4.83 ± 7.92	3.78 ± 7.79	(0.044)**
FBG (mg/dl)	119.16 ± 62.75	90.95 ± 35.54	(0.001)**
Anthropometric measures			
PAL	1.52 ± 0.245	1.58 ±0.28	(<0.001)**
BMI (kg/$\boldsymbol{m^2}$)	28.83 ± 4.56	27.77 ± 4.67	(<0.001)**
Age (year)	54.3 ± 6.86	48.29 ± 4.67	(<0.001)***
WC (cm)	98.9 ±10.7	95.2 ±11.87	(0.003)**
Blood pressure			
SBP (mm Hg)	133.89 ± 21.23	121.4± 17.71	(<0.001)**
DBP (mm Hg)	83.57 ± 10.84	79.12 ± 10.94	(<0.001)**

Data reported as mean ± standard deviation (SD).
Mann-Whitney U test.

Table 3. Descriptive statistics of qualitative clinical and biochemical characteristics of the study population.

Variable	Levels	N (CAD +) (%)	N (CAD -) (%)	*P*-value
Gender	Male	124 (52.8%)	3142 (44%)	0.321
	Female	111(47.2%)	4007 (56%)	
Marital status	Married	217(92.3%)	6718 (94%)	0.308
	Single	18 (7.7%)	432 (6%)	
Smoking	Yes	141 (60%)	4945 (69.2%)	0.003**
	No	94 (40%)	2205 (30.8%)	
Family History	Yes	131 (55.7%)	4728 (66.1%)	(<0.001)**
	No	104 (44.3%)	2422 (33.9%)	
Educational attainment	Low	154 (65.5%)	3781(52.9%)	(<0.001)**
	Moderate	60 (25.5%)	2514 (35.2%)	
	High	21 (8.9%)	855 (12%)	
Depression	Minimal	123 (52.3%)	4632 (64.8%)	
	Mild	50 (21.3%)	1219 (17%)	(<0.001)**
	Moderate	47 (20%)	874 (12.2%)	
	Severe	15 (6.4%)	425 (5.9%)	
Anxiety	Minimal	92 (39.1%)	3692 (51.6%)	
	Mild	70 (29.8%)	1916 (26.8%)	(<0.001)**
	Moderate	41 (17.4%)	976 (13.7%)	
	Severe	32 (13.6%)	566 (7.9%)	
	student	0 (0%)	13 (0.2%)	
Job	Employment	70 (29.8%)	28644 (40.1%)	(<0.001)**
	Unemployment	119 (50.6%)	3543 (49.6%)	
	Retiered	46 (19.6%)	730 (10.2%)	

*** *P*-value < 0.001; ** *P* -value < 0.05; * *P* -value < 0.2; Chi Square test (χ^2).

[DOI: 10.47176/mjiri.37.46]

Table 4. Predictive accuracies of discriminative approaches

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)	MSE
ANFIS	80	86	83.4	0.166
LR	70	74	72.4	0.175
FDA/MARS	74	81	77.7	0.223

Table 5. Confidence interval for sensitivity & specificity of discriminative approaches

Methods		Point Estimate	Lower confidence Interval	Upper confidence Interval
ANFIS	Sensitivity	0.8	0.77706	0.82938
	Specificity	0.86	0.83308	0.88588
LR	Sensitivity	0.7	0.67032	0.71188
	Specificity	0.74	0.70383	0.76268
FDA/MARS	Sensitivity	0.74	0.71395	0.76195
	Specificity	0.81	0.79362	0.83075

shown in Figure 3.

Discussion

Cardiovascular disease is the leading cause of death worldwide, according to the World Health Organization(3). This disease imposes huge health, social and economic burdens on society and often in affects people in the productive age group (25).

With the important role of correct and early diagnosis of disease as well as the expensive and high-risk diagnostic methods, using machine learning techniques and data mining algorithms in the diagnosis of cardiovascular disease has become of critical importance.

Since medical data analysis is highly sensitive and misdiagnosis will lead to irreparable errors, it is important to use the most accurate method with the highest accuracy and least error to analyze the data.

In the present cross-sectional study, three different methods, adaptive neuro-fuzzy inference system (ANFIS), logistic regression, and discriminative analysis, were used to diagnose cardiovascular disease in people aged over 35 years. Our findings showed that ANFIS had an 83.4% accuracy rate, 83.4% AUC, and 80% sensitivity with the highest ability in accurate medical decision-making for the diagnosis of cardiovascular disease. While logistic regression and discriminative analysis had an accuracy of about 74%, the ability of the ANFIS model in classification has also been proven in similar studies.

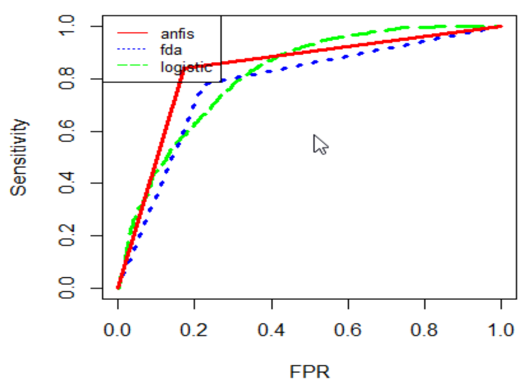

Figure 2. ROC curves for discriminative techniques

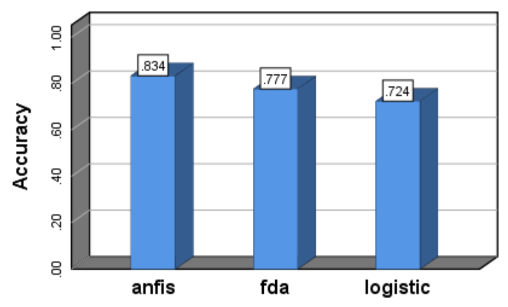

Figure 3. Comparison accuracy of ANFIS, LR and DA models

Bhuvaneswari et al., in the study on the heart disease dataset, used Principal Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict the risk of cardiovascular disease. The classification accuracy in their study was 93.2% (26). In 2014, Ziasabounchi et al., applied the ANFIS method to classify the degree of heart disease in 303 patients. Their results showed a 92.3% accuracy rate for forecasting a patient's degree of heart disease(8). Sagir et al., in a study to predict heart disease, evaluated the classification of two discrete Adaptive Neuro-Fuzzy Inference System models, ANFIS Matlab's built-in model (ANFIS_ LSGD) and a newly ANFIS model with Levenberg-Marquardt algorithm. Outputs illustrated the effectiveness of the ANFIS performance (27). Zabbah et al., in an analytical study on 200 subjects of cardiovascular patients, implemented an artificial neural network and neuro-fuzzy method to model the diagnosis of coronary artery disease. Implementation of neural network algorithms and fuzzy logic showed that although neural networks alone have a significant ability to diagnose coronary artery disease, combining neural networks and fuzzy logic incredibly increased the system's ability to detect the disease. The MSE for the artificial neural network and neuro-fuzzy method were 0.25 and 0.007, respectively (9). It seems that using neural network methods in the diagnosis of coronary heart disease appears effective.

The ANFIS model has also been used to diagnose other diseases. In Polat's study, an adaptive neuro-fuzzy inference system (ANFIS) and linear and quadratic discriminant analysis were used to detect diabetes. The classification accuracy for the ANFIS method obtained was 89.47 % (13). Mahmoudi et al. used the ANFIS method for cancer classification on microarray gene expression data. They compared ANFIS with Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Classification and Regression Trees (CART). ANFIS was the best model (14). Sancar et al., in their study, examined ANFIS performance

[DOI: 10.47176/mjiri.37.46]

to estimate Body Mass Index (BMI). Low Root Mean Square Error (RMSE) results of the model indicated that this model is biologically accepted and could be used for predicting BMI (28).

Conclusion

In conclusion, our study showed that the ANFIS method had the highest accuracy for detecting CAD and could be used in the classification of cardiovascular disease.

Acknowledgments

This study was supported by the Student Research Committee of Mashhad University of Medical Sciences, Mashhad, Iran (Project No. 981122).

Conflict of Interests

The authors declare that they have no competing interests.

References

1. Sadeghi M, Haghdoost AA, Bahrampour A, Dehghani M. Modeling the burden of cardiovascular diseases in Iran from 2005 to 2025: the impact of demographic changes. Iran J Public Health. 2017;46(4):506.
2. Forouzanfar MH, Sepanlou SG, Shahraz S, Dicker D, Naghavi P, Pourmalek F, et al. Evaluating causes of death and morbidity in Iran, global burden of diseases, injuries, and risk factors study 2010. Arch Iran Med. 2014;17(5):0-.
3. World Health Organization. Cardiovascular disease [Internet]. Geneva; 2019. https://www.who.int/cardiovascular_diseases/en.
4. Kaur K, Singh M. Heart disease prediction system using ANOVA, PCA and SVM classification. Int J Adv Res Innov Ideas Educ. 2016;2(3):1-6.
5. Wiener C, Fauci A, Braunwald E, Kasper D, Hauser S, Longo D, et al. Harrisons Principles of Internal Medicine Self-Assessment and Board Review 18th Edition. 2012.
6. Crawford MH, Education M-H. Current diagnosis & treatment in cardiology: McGraw Hill Medical; 2009.
7. Miao KH, Miao JH, Miao GJ. Diagnosing coronary heart disease using ensemble machine learning. Int J Adv Comput Sci Appl. 2016;7(10).
8. Ziasabounchi N, Askerzade I. ANFIS based classification model for heart disease prediction. Int J Comput Sci. 2014;14(02):7-12.
9. Zabbah I, Koohjani Z, Maroosi A, Layeghi K. Diagnosis of Coronary Artery Disease using Neuro-fuzzy-based method. . J Torbat Heydariyeh Univ Med Sci. 2018;6(3):48-59.
10. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl. 2008;34(1):366-74.
11. Persi Pamela I, Gayathri P, Jaisankar N. A fuzzy optimization technique for the prediction of coronary heart disease using decision tree. Int J Eng. 2013;5(3):2506-14.
12. Tahamtan RAM, Esmaeili MH, Ghaemian A, Esmaeili J. Application of artificial neural network for assessing coronary artery disease. J Maz Univ Med Sci. 2012;22(86):8-17.
13. Polat K, Güneş S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digit Signal Process. 2007;17(4):702-10.
14. Mahmoudi S, Lahijan BS, Kanan HR, editors. ANFIS-based wrapper model gene selection for cancer classification on microarray gene expression data. 2013 13th Iranian Conference on Fuzzy Systems (IFSC); 2013: IEEE.
15. Xu H, Duan Z, Miao C, Geng S, Jin Y. Development of a diagnosis model for coronary artery disease. Indian Heart J. 2017;69(5):634-9.
16. Gao T. Hybrid classification approach of SMOTE and instance selection for imbalanced datasets. Iowa State University. 2015.
17. Kirisci M, Yilmaz H, Saka MU. An ANFIS perspective for the diagnosis of type II diabetes. Ann Fuzzy Math. 2019;17(2):101-13.
18. Wu JD, Hsu CC, Chen HC. An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Syst Appl. 2009;36(4):7809-17.
19. Garg VK, Bansal R, editors. Soft computing technique based on ANFIS for the early detection of sleep disorders. 2015 International Conference on Advances in Computer Engineering and Applications; 2015: IEEE.
20. Zurada JM. 27. Introduction to Artificial Neural Systems. Jaico Publishing House. Mumbai, 1999; 121,.
21. Haykin S. Neural networks: a comprehensive foundation: Prentice Hall PTR; 1998.
22. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
23. Agresti A, Kateri M. Categorical data analysis (pp. 206-208). SpringerBerlin Heidelberg. 2011.
24. Hastie T, Tibshirani R, Buja A. Flexible discriminant analysis by optimal scoring. J Am Stat Assoc. 1994;89(428):1255-70.
25. Hadaegh F, Harati H, Ghanbarian A, Azizi F. Prevalence of coronary heart disease among Tehran adults: Tehran Lipid and Glucose Study. East Mediterr Health J. 2009;15(1):157-166.
26. NG BA, editor An intelligent approach based on Principal Component Analysis and Adaptive Neuro Fuzzy Inference System for predicting the risk of cardiovascular diseases. 2013 Fifth International Conference on Advanced Computing (ICoAC); 2013: IEEE.
27. Sagir AM, Sathasivam S. A Novel Adaptive Neuro Fuzzy Inference System Based Classification Model for Heart Disease Prediction. J Sci Technol. 2017;25(1).
28. Sancar N, Tabrizi SS. Body mass index estimation by using an adaptive neuro fuzzy inference system. Procedia Comput Sci. 2017;108:2501-6.