Expert Systems with Applications
Volume 34, Issue 1, January 2008, Pages 366-374

Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease

Imran Kurt [a], Mevlut Ture [b], A. Turhan Kurum [c]

https://doi.org/10.1016/j.eswa.2006.09.004
Get rights and content

Abstract

In this study, performances of classification techniques were compared in order to predict the presence of coronary artery disease (CAD). A retrospective analysis was performed in 1245 subjects (865 presence of CAD and 380 absence of CAD). We compared performances of logistic regression (LR), classification and regression tree (CART), multi-layer perceptron (MLP), radial basis function (RBF), and self-organizing feature maps (SOFM). Predictor variables were age, sex, family history of CAD, smoking status, diabetes mellitus, systemic hypertension, hypercholesterolemia, and body mass index (BMI). Performances of classification techniques were compared using ROC curve, Hierarchical Cluster Analysis (HCA), and Multidimensional Scaling (MDS). Areas under the ROC curves are 0.783, 0.753, 0.745, 0.721, and 0.675, respectively for MLP, LR, CART, RBF, and SOFM. MLP was found the best technique to predict presence of CAD in this data set, given its good classificatory performance. MLP, CART, LR, and RBF performed better than SOFM in predicting CAD in according to HCA and MDS.

Keywords

Logistic regression; Decision tree; Neural networks; Coronary artery disease; Multidimensional scaling; Hierarchical cluster analysis; ROC curve

1. Introduction

Coronary artery disease (CAD) is a major worldwide health problem with its incidence and mortality rates (Backer et al., 2003). Many risk factors are known to play role in pathogenesis of CAD and myocardial infarction. Family history, smoking, hypertension, hypercholesterolemia, diabetes mellitus and obesity have been described as the major risk factors for CAD (Burke et al., 1997, Celermajer et al., 1993, Chobanian et al., 2003, Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001, Fuster, 1994, Haskell et al., 1994, Herrington et al., 2003, Higgins et al., 1992, Rahman et al., 2004, The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 2003, Zeiher et al., 1993). Identification of risk factors in CAD is essential for the management and follow-up of CAD. Numerous cross-sectional angiography studies have reported correlations of one or more major risk factors for myocardial infarction with the presence of CAD.

Predicting the outcome of a disease is one of the most interesting and challenging tasks in which to develop data mining applications. Some applications of neural networks in cardiology were performed for the analysis of heart sounds, analysis of cardiac arrhythmias, the detection of ventricular ectopic activity, and the detection of atrial fibrillation. Furthermore, the development of implantable devices for treatment of life-threatening arrhythmias has simulated intracardiac rhythm classification using neural networks. Neural networks have been trained to recognize ST-T segment changes, to recognize CAD in general, to predict the number of vessels involved and to identify three-vessel and mainstem disease, even at rest (Dassen, Egmont-Petersen, & Mulleneers, 1998). Few works have been published on the comparison of classification techniques in different areas. Moisen and Frescino (2002) compared linear models, generalized additive models, classification and regression tree (CART), Multivariate Additive Regression Splines (MARS), and artificial neural networks for mapping forest characteristics in the Interior Western United States using forest inventory field data and ancillary satellite-based information. Ture, Kurt, Kurum, and Ozdamar (2005) compared various classification techniques to predict control and hypertension groups. They created models using logistic regression (LR), flexible discriminant analysis (FDA), FDA with MARS (FDA/MARS), chi-squared automatic interaction detector (CHAID), quick unbiased efficient statistical tree (QUEST), CART, radial basis function (RBF) and multi-layer perceptron (MLP) to predict control and hypertension groups. Delen, Walker, and Kadam (2004) compared LR, decision tree (C5) and artificial neural networks for predicting the survivability of diagnosed cases for breast cancer. Stark and Pfeiffer (1999) compared LR, classification tree algorithms (ID3, C4.5, CHAID, CART) and artificial neural networks to solve classification problems in complex data sets in

veterinary epidemiology. Colombet et al. (2000) evaluated the implementation and performance of CART and artificial neural networks comparatively with a LR model, in order to predict the risk of cardiovascular disease in a real database. King, Feng, and Sutherland (1995) compared symbolic learning (CART, C4.5, NewID, AC^2, ITrule, Cal5, and CN2), statistics (Naı¨ve Bayes, *k*-nearest neighbor, kernel density, linear discriminant, quadratic discriminant, LR, projection pursuit, and Bayesian networks), and neural networks (back-propagation and RBF) algorithms on twelve datasets with respect to large real-world problems.

The purpose of this study is to compare performances of classification techniques in order to predict the presence of CAD. We have created models using LR, CART, neural networks algorithms (RBF, MLP, and self-organizing feature maps (SOFM)) that they are often used for classification problems. LR is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of set of independent variables which are continuous, categorical, or both. Furthermore, it assumes that measures of dependent variables are independently and randomly sampled, all potentially relevant independent variables are in the model and all independent variables in the model are relevant (Hosmer and Lemeshow, 2000, Kleinbaum, 1994). CART is inherently non-parametric that no assumptions are made regarding the underlying distribution of values of the predictor variables. Thus, CART can handle numerical data that are highly skewed or multi-modal, as well as categorical predictors with either ordinal or non-ordinal structure (Breiman, Friedman, Olshen, & Stone, 1984). Neural networks have been used to model medical and functional outcomes of dangerous disease. They have become a popular tool for classification, as they are very flexible, not assuming any parametric form for distinguishing between categories (Lee, 2001).

Performances of classification techniques were compared using ROC curve, Hierarchical Cluster Analysis, and Multidimensional Scaling.

2. Material and methods

2.1. Data

A retrospective analysis was performed in 1245 subjects (865 presence of CAD and 380 absence of CAD). Clinically relevant CAD was defined by the presence of at least one vessel with a stenosis $\geqslant$50% in coronary angiography to be done due to angina associated with evidence for myocardial ischemia either by stress electrocardiography, stress Tc99 MIBI scintigraphy or pathologic resting electrocardiography. All angiograms were assessed by two cardiologists not participating in the study. All patients with suspected CAD were seen in the Cardiology Clinic of Trakya University Medical Faculty in Turkey between January 2002 and February 2003. The patients had coronary angiography because of stable angina pectoris or an acute coronary syndrome or atypical angina was included in the study. Patients with non-atherosclerotic CAD were excluded from the study.

Independent variables included age, sex, family history of CAD, smoking status, diabetes mellitus, systemic hypertension, hypercholesterolemia, and BMI. Hypertension was diagnosed when the systolic blood pressure (BP) was ⩾140 mmHg and/or diastolic BP was ⩾90 mmHg on at least three separate occasions and was established by the absence of clinical findings suggestive of secondary form of hypertension (Chobanian et al., 2003). BP was measured in the sitting position in a quite room, using a mercury sphygmomanometer, after the patient had rested for at least 10 min. Systolic BP was recorded at the appearance of sounds (korotkoff phase I) and the diastolic at their disappearance (korotkoff phase V). Patients who currently smoked or discontinued during the last 6 months were categorized as smokers. Diabetes was considered confirmed if the questionnaire indicated one of the following National Diabetes Data Group criteria: (1) Symptoms of diabetes plus casual plasma glucose concentration ⩾200 mg/dl (11.1 mmol/l). Casual is defined as any time of day without regard to time since last meal. The classic symptoms of diabetes include polyuria, polydipsia, and unexplained weight loss. (2) FPG⩾126 mg/dl (7.0 mmol/l). Fasting is defined as no caloric intake for at least 8 h. (3) 2-h PG⩾200 mg/dl (11.1 mmol/l) during an OGTT. The test should be performed as described by WHO hypertension (Chobanian et al., 2003), using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water (The Expert Committee on the Diagnosis & Classification of Diabetes Mellitus, 2003). Hypercholesterolemia was accepted when the total cholesterol was above 200 mg/dl (Expert Panel on Detection, Evaluation, & Treatment of High Blood Cholesterol in Adults, 2001). Measurement of routine laboratory examinations of serum and urine were performed in the fasting state.

Before building models, the data set were randomly split into three subsets, 60% (n=747) of the data for training set (for NN, LR and CART), 20% (n=249) of data for test set (for NN, LR and CART) and 20% (n=249) of data for cross validation set (for NN).

2.2. Classification techniques

2.2.1. Logistic regression analysis

LR is a regression method for predicting a dichotomous dependent variable. In producing the LR equation, the maximum-likelihood ratio was used to determine the statistical significance of the variables (Hosmer and Lemeshow, 2000, Ozdamar, 2004).

LR is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous.

LR model for p independent variables can be written as

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p)}}$$

where $P(Y=1)$ is probability of presence of CAD, and $\beta_0,\beta_1,\ldots,\beta_p$ are regression coefficients. There is

a linear model hidden within the logistic regression model. The natural logarithm of the ratio of $P(Y=1)$ to $(1-P(Y=1))$ gives a linear model in X_i:

$$g(x) = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

The $g(x)$, has many of the desirable properties of a linear regression model. The independent variables can be a combination of continuous and categorical variables (Hosmer and Lemeshow, 2000, Ozdamar, 2004).

LR model can include the main effects and interaction terms. An important step in the process of modeling a set of data is determining whether there is evidence of interaction and confounder term in the data. The term confounder is used to describe a covariate that is associated with both the dependent variable of interest and a primary independent variable. When both associations are present then the relationship between independent variable and the dependent variable is said to be confounded. LR model to check for the confounder status of a covariate is to compare the estimated coefficient for the independent variable from models containing and not containing the covariate. Any clinically important change in the estimated coefficient for the independent variable suggests that the covariate is a confounder and should be included in the model, regardless of the statistical significance of its estimated coefficient. One way to test for confounder and interactions in LR is to start with a main effects model, and use a forward-selection method to find interaction terms which significantly reduce the likelihood ratio test statistic (Hosmer & Lemeshow, 2000).

In our study, we used LR with a forward stepwise variable selection to predict the probability of presence of CAD. We created the best model included main effects and interaction terms using the likelihood ratio test. Fit was assessed by chi-square statistics proposed by Hosmer–Lemeshow Goodness of Fit Test.

2.2.2. Classification and regression tree

CART is a recursive partitioning method to be used both for regression and classification. CART is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of impurity or diversity measures (Gini, twoing, ordered twoing and least-squared deviation). The goal is to produce subsets of the data which are as homogeneous as possible with respect to the target variable (Breiman et al., 1984). In this study, we used measure of Gini impurity that used for categorical target variables.

Gini Impurity Measure:

The Gini index at node t, $g(t)$, is defined as

$$g(t) = \sum_{j \neq i} p(j|t)p(i|t)$$

where i and j are categories of the target variable. The equation for the Gini index can also be written as

$$g(t) = 1 - \sum_j p^2(j|t)$$

Thus, when the cases in a node are evenly distributed across the categories, the Gini index takes its maximum value of $1-(1/k)$, where k is the number of categories for the target variable. When all cases in the node belong to the same category, the Gini index equals 0.

If costs of misclassification are specified, the Gini index is computed as

$$g(t) = \sum_{j \neq i} C(i|j)p(j|t)p(i|t)$$

where $C(i|j)$ is the probability of misclassifying a category j case as category i.

The Gini criterion function for split s at node t is defined as

$$\Phi(s,t) = g(t) - p_{\mathrm{L}}g(t_{\mathrm{L}}) - p_{\mathrm{R}}g(t_{\mathrm{R}})$$

where p_{L} is the proportion of cases in t sent to the left child node, and p_{R} is the proportion sent to the right child node. The split s is chosen to maximize the value of $\Phi(s,t)$. This value is reported as the improvement in the tree (Breiman et al., 1984).

2.2.3. Multi-layer perceptron

MLP is feed-forward neural networks trained with the standard back-propagation algorithm. It is supervised networks so they require a desired response to be trained. It learns how to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input–output map. It has been shown to approximate the performance of optimal statistical classifiers in difficult problems. The most popular static network is the MLP (Haykin, 1999, Principe et al., 1999).

The MLP is trained with error correction learning, which is appropriate here because the desired MLP response is the arteriographic result and as such known. Error correction learning works in the following way: From the system response at neuron j at iteration t, $y_j(t)$, and the desired response $d_j(t)$ for a given input pattern an instantaneous error $e_j(t)$ is defined by $e_j(t)=d_j(t)-y_j(t)$.

Using the theory of gradient descent learning, each weight in the network can be adapted by correcting the present value of the weight with a term that is proportional to the present input and error at the weight, i.e.

$$w_{jk}(t+1) = w_{jk}(t) + \eta\delta_j(t)x_k(t)$$

The $\eta(t)$ is the learning-rate parameter. The $w_{jk}(t)$ is the weight connecting the output of neuron k to

the input neuron j at iteration t. The local error $\delta_j(t)$ can be directly computed from $e_j(t)$ at the output neuron or can be computed as a weighted sum of errors at the internal neurons.

Back-propagation computes the sensitivity of a cost functional with respect to each weight in the network, and updates each weight proportional to the sensitivity. The beauty of the procedure is that it can be implemented with local information and requires just a few multiplications per weight, which is very efficient. Because this is a gradient descent procedure, it only uses the local information so can be caught in local minima. Moreover, the procedure is inherently noisy since we are using a poor estimate of the gradient, causing slow convergence.

Momentum learning is an improvement to the straight gradient descent in the sense that a memory term (the past increment to the weight) is used to speed up and stabilize convergence. In momentum learning the equation to update the weights becomes

$$w_{jk}(t+1) = w_{jk}(t) + \eta\delta_j(t)x_k(t) + \alpha(w_{jk}(t) - w_{jk}(t-1))$$

where α is the momentum. Normally α should be set between 0.1 and 0.9 (Hansen and Salamon, 1990, Haykin, 1999, Principe et al., 1999).

2.2.4. Radial basis function

RBF networks have a static Gaussian (bell-shaped) function as the nonlinearity for the hidden layer processing elements. The Gaussian function responds only to a small region of the input space where the Gaussian is centered. The key to a successful implementation of these networks is to find suitable centers for the Gaussian functions. This can be done with supervised learning, but an unsupervised approach usually produces better results. The advantage of the radial basis function network is that it finds the input to output map using local approximators. Usually the supervised segment is simply a linear combination of the approximators. Since linear combiners have few weights, these networks train extremely fast and require fewer training samples (Haykin, 1999, Principe et al., 1999).

Networks can be constructed in the following way: Bring every input component (p) to a layer of hidden neurons. Each node in the hidden layer is a p multivariate Gaussian function

$$G(x;C_r) = \mathrm{e}^{-\left(\frac{\|x-C_r\|}{2\sigma_r^2}\right)^2} \quad (r = 1,2,\ldots,k)$$

of mean C_r (each data point) and σ_r variance. These functions are called radial basis functions. The k is the number of neurons at hidden layer. Finally, linearly weight the output of the hidden neuron to obtain

$$F(x) = \sum_{r=1}^{k} w_r G(x;C_r)$$

The problem with this solution is that it may lead to a very large hidden layer (Haykin, 1999, Principe et al., 1999).

2.2.5. Self-organizing feature maps

SOFM transform the input of arbitrary dimension into a one or two dimensional discrete map subject to a topological (neighborhood preserving) constraint. The feature maps are computed using Kohonen unsupervised learning. SOFM network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input. The output of the SOFM can be used as input to a supervised classification neural network such as the MLP. This network's key advantage is the clustering produced by the SOFM which reduces the input space into representative features using a self-organizing process. Hence the underlying structure of the input space is kept, while the dimensionality of the space is reduced (Haykin, 1999, Principe et al., 1999).

The SOFM layer can be a one or two dimensional lattice, and the size of the net provides the resolution for the lattice. The SOFM algorithm is as follows: Initialize the weights with small different random values for symmetry breaking.

For each input data, SOFM find the best-matching (winning) neuron ℓ at time step t by using the minimum-distance Euclidean criterion:

$$\ell = \arg\min_j \|x(t) - w_j\|$$

For the winning neuron, it adjusts the weight vectors of all neurons by using the update formula:

$$w_j(t+1) = w_j(t) + \eta(t)\Lambda_{j,\ell}(t)[x(t) - w_j(t)]$$

where $w_j(t)$ is the weight of neuron j, and $\Lambda_{j,\ell}(t)$ is the neighborhood function centered around the winning neuron ℓ; both $\eta(t)$ and $\Lambda_{j,\ell}(t)$ are varied dynamically learning for best results (Haykin, 1999, Principe et al., 1999).

2.3. Hierarchical cluster analysis (HCA)

HCA is a statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. When there are N cases, this involves $N-1$ clustering steps, or fusions. This hierarchical clustering process can be represented as a tree, or dendrogram, where each step in the clustering process is illustrated by a join of the tree (Sharma, 1996).

2.4. Multidimensional scaling (MDS)

MDS is a method that represents measurements of similarity (or dissimilarity) among pairs of objects as distances between points of a low-dimensional space. It helps us to represent

(dis)similarities between objects as distances in a Euclidean space. In effect, the more dissimilar two objects are, the larger the distance between the objects in the Euclidean space should be. The objects in our study are the different classification techniques described by their characteristics in terms of classification performance measurements. The location of the techniques on the map is based on their position in the m-dimensional variable space. Similar to the R^2 measure in regression analysis, there is a pseudo-R^2 calculated in MDS. The pseudo-R^2 is equal to the percentage of the sum of squared dissimilarities explained by the model. Another goodness-of-fit measure of the projection is the so-called stress factor. A stress factor <0.05 is considered to be good (Borg and Groenen, 1997, Kruskal, 1964).

3. Results

3.1. Comparison of characteristics

The characteristics of the study population are shown in Table 1. We performed the classical statistical analysis to examine the difference in the distribution of age variable between the presence and absence of CAD. Age was tested for normal distribution by the Shapiro–Wilk test. Comparison between two groups was made by the Mann–Whitney U-test for non-normally distributed age variable. Nominal variables were tested by chi-square test for presence and absence of CAD.

Table 1. Characteristics of study subjects

Characteristics	**CAD**		***P*-value**
	Presence (*n*=865)	**Absence (*n*=380)**	
Sex (male/female)	4.1	1.1	<0.001
Age (years)	60 (52–67)	55 (45–64)	<0.001
BMI (%)			0.024
Thin	0.8	2.6	
Normal	30.9	27.6	
Obese	68.3	69.7	
Smoking Status (%)	57.9	42.4	<0.001
Diabetes Mellitus (%)	23.6	9.2	<0.001
Systemic hypertension (%)	47.1	41.6	0.074
Hypercholesterolemia (%)	44.7	41.3	0.262
Family history of CAD (%)	25.2	21.3	0.139

Age variable is median (25–75% percentiles); Smoking status: current or discontinued during the last 6 months; Diabetes mellitus: HbA1c ⩾6.2%, current insulin, or oral hypoglicemic therapy; Systemic hypertension: >140/90 mmHg or current antihypertensive therapy; Hypercholesterolemia: cholesterol ⩾200 mg/dl or current lipid-lowering drugs; CAD: coronary artery disease.

Five variables (sex, age, BMI, smoking status and diabetes mellitus) were independently significant between presence and absence of CAD (respectively $P<0.001$, $P<0.001$, $P<0.05$, $P<0.001$ and $P<0.001$).

Using the independent predictors, we used LR analysis with a forward stepwise variable selection to predict the probability of presence of CAD. We created the best model included main effects and interaction terms. Fit was assessed by chi-square statistics proposed by Hosmer–Lemeshow Goodness-of-Fit Test, and LR model showed a good fit based on the chi-square of Hosmer–Lemeshow goodness of fit statistics ($P>0.05$). We found logit model by LR analysis:

$$P(\text{CAD present}) = \frac{1}{1+e^{-3.874-1.639\ \text{Sex(F)}+0.054\text{Age}+0.927\text{DM}+1.684\ \text{BMI(N)}+2.082\ \text{BMI(O)}+1.733\ \text{BMI(N)}\ \text{DM}}}$$

where F is female, N is normal, O is obese, and DM is diabetes mellitus.

3.2. Comparison of classification techniques

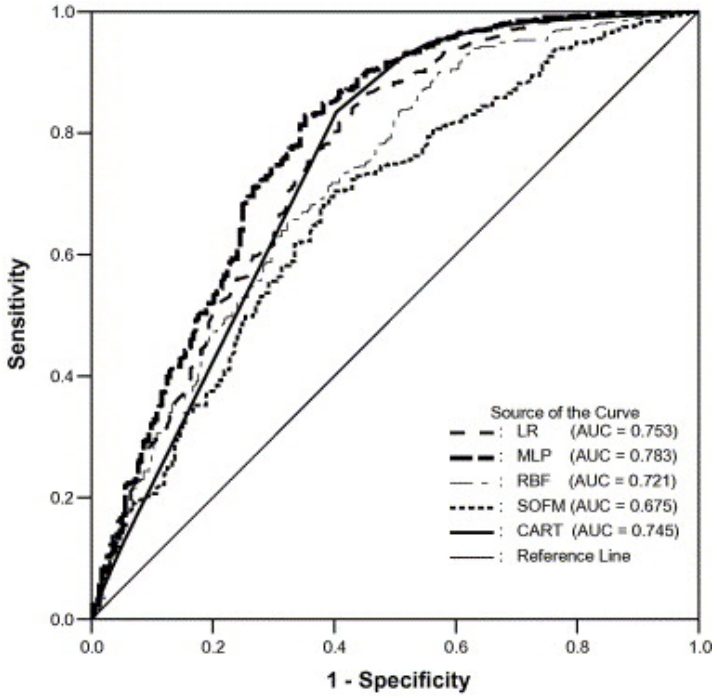
A comparison of the sensitivity (SEN), specificity (SPE), positive predictive rate (PPR), negative predictive rate (NPR) and correct classification frequency (CCF) for training set and test set of classification techniques are shown in Table 2.

Table 2. Comparison of the performance of models for training set and test set

Model	Training set					Test set				
	SEN (%)	SPE (%)	PPR (%)	NPR (%)	CCF (%)	SEN (%)	SPE (%)	PPR (%)	NPR (%)	CCF (%)
LR	91.8	43.3	78.1	70.6	76.7	92.3	45.6	81.9	68.9	79.5
CART	94.9	42.9	78.6	79.4	78.7	92.3	47.1	82.3	69.6	79.9
MLP	93.4	46.4	79.3	76.1	78.7	91.7	45.6	81.8	67.4	79.1
RBF	91.6	39.1	76.8	67.9	75.2	89.5	42.6	80.6	60.4	76.7
SOFM	97.9	9.9	70.5	67.6	70.4	98.9	7.4	74.0	71.4	73.9

SEN: Sensitivity, SPE: Specificity, PPR: Positive Predictive Rate, NPR: Negative Predictive Rate, CCF: Correct Classification Frequency.

LR, CART, MLP, RBF, and SOFM had an area under the Receiver Operating Characteristics (ROC) curve in the 0.675–0.783 range (Fig. 1). We tested the statistical significance of the difference between the areas under two ROC curves using the method proposed by Hanley and McNeil (1983) (Table 3). The performances of LR and MLP, LR and RBF, LR and SOFM were statistically different (P=0.001, P=0.015, and P<0.001, respectively). The performances of CART and MLP, CART and SOFM were statistically different (P=0.026, and P=0.003, respectively). The performances of MLP and RBF, MLP and SOFM, RBF and SOFM were statistically different (P<0.001, P<0.001, and P=0.013, respectively). However, the performances of LR and CART, CART and RBF was not statistically different (P=0.663 and P=0.250, respectively). As it can be seen from Fig. 1, MLP performed better than other techniques in predicting CAD. This technique was followed by LR and CART, and RBF showing the second and third best performance. The predictive performance of SOFM was lower than the performance of other models.



Download: Download full-size image

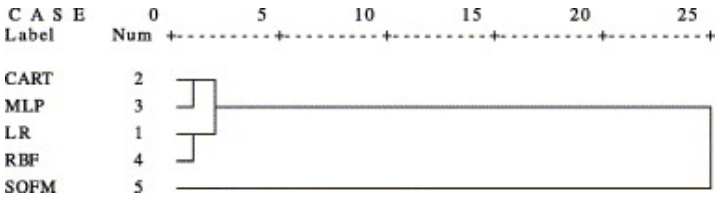
Fig. 1. ROC curves for classification techniques.

Table 3. Difference between areas under the ROC curves (AUC_d=AUC_{row}−AUC_{column}), standard errors (SE), and performances of classification techniques

	LR	CART	MLP	RBF
CART	AUC_d=−0.008 SE(AUC_d)=0.019 P=0.663			

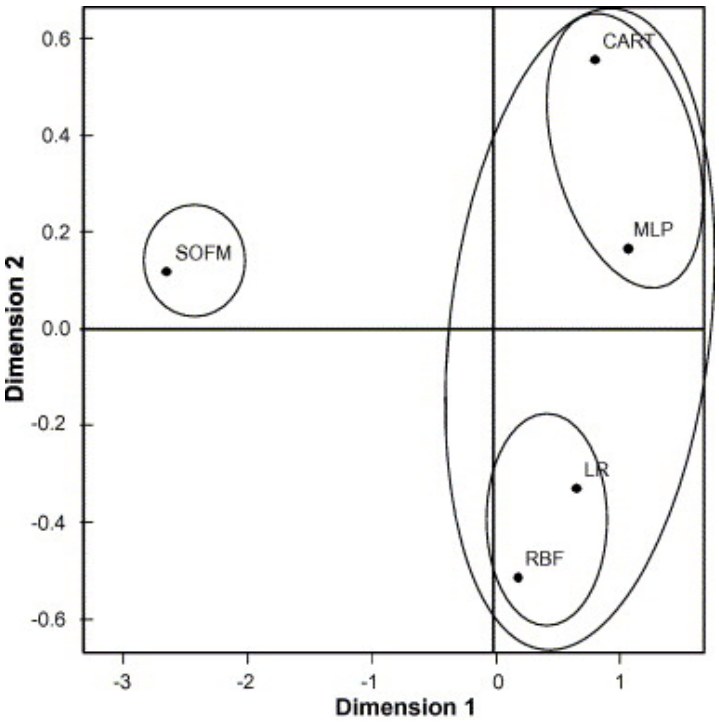
	LR	CART	MLP	RBF
MLP	AUC_d=0.030	AUC_d=0.038		
	SE(AUC_d)=0.009	SE(AUC_d)=0.017		
	P=0.001	*P*=0.026		
RBF	AUC_d=−0.032	AUC_d=−0.024	AUC_d=−0.062	
	SE(AUC_d)=0.013	SE(AUC_d)=0.021	SE(AUC_d)=0.014	
	P=0.015	*P*=0.250	*P*<0.001	
SOFM	AUC_d=−0.078	AUC_d=−0.070	AUC_d=−0.109	AUC_d=−0.046
	SE(AUC_d)=0.021	SE(AUC_d)=0.023	SE(AUC_d)=0.020	SE(AUC_d)=0.019
	P<0.001	*P*=0.003	*P*<0.001	*P*=0.013

SEN, SPE, PPR, NPR, and CCF for training set were used as input variables in both HCA and MDS. HCA and MDS were done to identify homogenous groups of classification techniques based on a SEN, SPE, PPR, NPR, and CCF. The dendrogram from between-groups linkage clustering method that was obtained is shown in Fig. 2. Five classification techniques were clustered into two distinct groups, the first group included MLP, CART, LR, and RBF and the second group included only SOFM. We found that the first dimension accounted for 98.7% and second dimension accounted for 1.3% of the variation and had a pseudo-R^2 of 1 (stress factor 0.00037) in MDS. These two-dimensions were plotted against each other in Fig. 3. MLP, CART, LR, and RBF techniques are located closest to each other but are located further than SOFM. As it can be seen from Fig. 2, Fig. 3, MLP, CART, LR, and RBF performed better than SOFM in predicting CAD.



Download: Download full-size image

Fig. 2. Dendrogram showing relationship among classification techniques.



Download: Download full-size image

Fig. 3. Euclidean distance model for classification techniques.

4. Discussion

Cardiovascular disease is the leading cause of mortality in Turkey as in rest of the world. The most important recent epidemiological investigation on Turkish population, the Turkish Adult Risk Factors Study, has shown that the incidence of CAD in young individuals is common in this population (Turkish Heart Report, 2000).

Cardiovascular disease is generally due to combination of several risk factors. Risk factor modifications have been unequivocally shown to reduce mortality and morbidity, especially in people with either unrecognized or recognized cardiovascular disease (Backer et al., 2003).

In this study, we report a research where we developed several prediction models for predicting CAD. Specially, we used LR, CART and three neural networks algorithms. Furthermore, we evaluated performance of models according to SEN, SPE, PPR, NPR, and PR. We tested the statistical significance of the difference between the areas under two ROC curves using the method proposed by Hanley and McNeil (1983).

Furthermore, HCA and MDS were done to identify homogenous groups of classification techniques based on the SEN, SPE, PPR, NPR, and PR.

According to previous studies, the method with the best classification performance may differ from one data structure to another. Moisen and Frescino (2002) found that generalized additive models

and MARS performed (marginally) better for prediction of forest characteristics than artificial neural networks, CART and linear models. Delen et al. (2004) found that decision tree (C5) performed the best than LR and artificial neural networks for predicting the survivability of diagnosed cases for breast cancer. Stark and Pfeiffer (1999) reported that classification trees (ID3, C4.5, CHAID and CART) were well-suited for exploratory data analysis in complex data sets in veterinary epidemiology. King et al. (1995) reported that there was not single best algorithm, although any comparative study is by its nature limited, a number of tentative conclusions could be drawn from within twelve datasets results with respect to image analysis, medicine, engineering, and finance, and the best algorithm for a particular data set depends crucially on features of that dataset. Colombet et al. (2000) found that the predictive performance of CART was slightly lower than the performance of LR and artificial neural networks. Ture et al. (2005) compared various classification techniques to predict control and hypertension groups. The latter group created models using LR, FDA, FDA/MARS, CHAID, QUEST, CART, RBF and MLP to predict control and hypertension groups and identified homogenous groups using HCA. They found that MLP and RBF performed better than other techniques. In the present study, we found that MLP, LR, CART, and RBF techniques performed better than SOFM. According to HCA and MDS, it was seen that MLP and CART performed (marginally) better for prediction of CAD than LR and RBF.

ROC curves are indispensable and conceptually straightforward, but selecting a single overall performance measure includes some problems. A lack of reporting of confidence intervals for measures of accuracy is a common problem in diagnostics. Other problems, in particular the dichotomization of a continuous-scale gold standard outcome, are specific to the kinds of diagnostic tests often evaluated by clinicians; further methodologic research is needed in these areas (Hilden, 2005, Obuchowski et al., 2004); the area under ROC curve is calculated merely based on SEN and SPE, but it does not use other performance measurements. Researchers should not prefer quickly to examination of the area under the ROC curve without pausing to ask whether the ROC curve is the right, or the best, measure. On the other hand, MDS is a technique for visual assessment of the similarities between classification techniques by creating a map using typically 2–3 dimensions, and also HCA is a technique for dendrogram assessment of the distances between classification techniques using all performance measurements. Hence, HCA and MDS are useful methods for comparison of the performances of models.

For neural networks applications in cardiology, Dassen et al. (1998) reported that a major advantage of using neural networks to model the relationship between the possible signs and symptoms and the diagnosis is the fact that this relationship does not have to be a linear one. However, despite the wide interest in the application of neural networks there are a number of limitations that make the introduction of these tools in daily practice difficult. First, because of the black box nature of neural networks, it is difficult to explain. Second problem is how to validate a trained neural network. They suggested to cardiologists for avoiding from these problems. Similarly to this study, we suggest that the trained neural network should be evaluated using an independent test set; the need for a test set to evaluate criteria is also present for classical systems, but the ability to learn all cases by heart

makes it even more essential for neural networks; the most important advantages of neural networks are that it draws consistent conclusions and it can be built and evaluated using a large number of cases.

Each classification technique shows some characteristics which may be interesting in the context of clinical practice. First, the tree representation in CART is close to the medical reasoning and can help to structure the understanding of prediction. Second, model obtained with MLP is not fixed since the iterative learning process can continue on local data. These methods probably have the potential to complement existing statistical models and to contribute to the interpretation and presentation of risk in computerized decision support systems. These models provide a comprehensive analytic framework to inform the optimal design of clinical guidelines and health policy for the prevention and management of CAD.

We suggest that age, sex, family history of CAD, smoking status, systemic hypertension, diabetes mellitus, hypercholesterolemia, and BMI variables may be used as reliable indicators to predict presence of CAD. All models had not very high correct classification frequencies because our study had several limitations that were the lack of input variables for risk factors, such as exercise behaviour, lipoprotein (a), hyperuricemia, and homocysteinemia. In our study, we compared methods by using a real data set in order to provide information on general tendency of data structures in data sets and help researchers to select best method for solving problems of classification. There are limited data on sufficiency of classification efforts by only one method. On the basis of these considerations, we suggest that data should be better explored and processed by high performance modeling methods. Researchers should avoid assessment of data by using only one method in future studies focusing on CAD or any other clinical condition.

Acknowledgement

We would like to express our gratitude to Dr. Levent Ozturk for reviewing manuscript and for his useful comments.

Recommended articles

References

Backer et al., 2003 G.D. Backer, E. Ambrosioni, K. Borch-Johnsen, C. Brotons, R. Cifkova, J. Dallongeville, *et al.*

European guidelines on cardiovascular disease prevention in clinical practice. Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts)

European Heart Journal, 24 (2003), pp. 1601-1610

Google Scholar ↗

Borg and Groenen, 1997 I. Borg, P. Groenen
Modern multidimensional scaling theory and applications
Springer-Verlag, New York (1997)
Google Scholar ↗

Breiman et al., 1984 L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone
Classification and regression trees
Wadsworth and Brooks/Cole, Montery (1984)
Google Scholar ↗

Burke et al., 1997 A.P. Burke, A. Farb, G.T. Malcom, Y.H. Liang, J. Smialek, R. Virmani
Coronary risk factors and plaque morphology in men with coronary disease who died suddenly
New England Journal of Medicine, 336 (18) (1997), pp. 1276-1282
View in Scopus ↗ Google Scholar ↗

Celermajer et al., 1993 D.S. Celermajer, K.E. Sorensen, D. Georgakopoulos, C. Bull, O. Thomas, J. Robinson, *et al.*
Cigarette-smoking is associated with dose-related and potentially reversible impairment of endothelium-dependent dilation in healthy-young adults
Circulation, 88 (5) (1993), pp. 2149-2155
View in Scopus ↗ Google Scholar ↗

Chobanian et al., 2003 A.V. Chobanian, G.L. Bakris, H.R. Black, W.C. Cushman, L.A. Green, J.L. Izzo Jr., *et al.*
National Heart, Lung and Blood Institute Joint National Committee on Prevention, detection, evaluation, and treatment of high blood pressure; National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood pressure: the JNC 7 report
JAMA, 289 (2003), pp. 2560-2572
View in Scopus ↗ Google Scholar ↗

Colombet et al., 2000 I. Colombet, A. Ruelland, G. Chatellier, F. Gueyffier, P. Degoulet, M.C. Jaulent
Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression
Proc AMIA Symp (2000), pp. 156-160
View in Scopus ↗ Google Scholar ↗

Dassen et al., 1998 W.R.M. Dassen, M. Egmont-Petersen, R.G.A. Mulleneers
Artificial neural networks in cardiology; a review
P.E. Vardas (Ed.), Cardiac arrhythmias, pacing and electrophysiology, Kluwer Academic Publishers, Great Britain (1998), pp. 205-211

Crossref ↗ Google Scholar ↗

Delen et al., 2004 D. Delen, G. Walker, A. Kadam

Predicting breast cancer survivability: a comparison of three data mining methods

Artificial Intelligence in Medicine, 34 (2) (2004), pp. 113-127

Google Scholar ↗

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001 Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults

Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III)

JAMA, 285 (19) (2001), pp. 2486-2497

Google Scholar ↗

Fuster, 1994 V. Fuster

Mechanisms leading to myocardial-infarction-insights from studies of vascular biology

Circulation, 90 (4) (1994), pp. 2126-2146

View in Scopus ↗ Google Scholar ↗

Hanley and McNeil, 1983 J.A. Hanley, B.J. McNeil

A method of comparing the areas under receiver operating characteristic curves derived from the same cases

Radiology, 148 (3) (1983), pp. 839-843

Crossref ↗ View in Scopus ↗ Google Scholar ↗

Hansen and Salamon, 1990 L.K. Hansen, P. Salamon

Neural network ensembles

IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (10) (1990), pp. 993-1001

View in Scopus ↗ Google Scholar ↗

Haskell et al., 1994 W.I. Haskell, E.L. Alderman, J.M. Fair, D.J. Maron, S.F. Mackey, H.R. Superko, *et al.*

Effects of intensive multiple risk factor reduction on coronary atherosclerosis and clinical cardiac events in men and women with coronary—artery disease—the Stanford-coronary-risk-intervention-project

Circulation, 89 (3) (1994), pp. 975-990

View in Scopus ↗ Google Scholar ↗

Haykin, 1999 S. Haykin

Neural network: A comprehensive foundation

Prentice Hall, Upper Saddle River, NJ (1999)

Google Scholar ↗

Herrington et al., 2003 D.M. Herrington, K. Kesler, J.C. Reiber, W. Davis, W.V. Brown, R. Helms, *et al.*
Arterial compliance adds to conventional risk factors for prediction of angiographic coronary artery disease
American Heart Journal, 146 (2003), pp. 662-667
View PDF View article View in Scopus ↗ Google Scholar ↗

Higgins et al., 1992 T.I. Higgins, F.G. Estafanous, F.D. Loop, G.J. Beck, J.M. Blum, L. Paranandi
Stratification of morbidity and mortality outcome by preoperative risk-factors in coronary–artery bypass patients—a clinical severity score
JAMA, 267 (17) (1992), pp. 2344-2348
Crossref ↗ View in Scopus ↗ Google Scholar ↗

Hilden, 2005 J. Hilden
What properties should an overall measure of test performance possess?
Clinical Chemistry, 51 (2) (2005), pp. 471-472
Crossref ↗ View in Scopus ↗ Google Scholar ↗

Hosmer and Lemeshow, 2000 D.W. Hosmer, S. Lemeshow
Applied logistic regression
John Wiley & Sons, New York (2000)
Google Scholar ↗

King et al., 1995 R.D. King, C. Feng, A. Sutherland
Statlog-comparison of classification algorithms on large real-world problems
Applied Artificial Intelligence, 9 (3) (1995), pp. 289-333
Crossref ↗ View in Scopus ↗ Google Scholar ↗

Kleinbaum, 1994 D.G. Kleinbaum
Logistic regression: A self-learning text
Springer-Verlag, New York (1994)
Google Scholar ↗

Kruskal, 1964 J.B. Kruskal
Multidimensional scaling by optimizing goodness of fit to nonmetric hypothesis
Psychometrika, 29 (1964), pp. 1-27
View in Scopus ↗ Google Scholar ↗

Lee, 2001 H.K.H. Lee
Model Selection for neural network classification
Journal of Classification, 18 (2001), pp. 227-243
Crossref ↗ View in Scopus ↗ Google Scholar ↗

Moisen and Frescino, 2002 G.G. Moisen, T.S. Frescino

Comparing five modelling techniques for predicting forests characteristics

Ecological Modelling, 157 (2002), pp. 209-225

View PDF View article View in Scopus ↗ Google Scholar ↗

Obuchowski et al., 2004 N.A. Obuchowski, M.L. Lieber, F.H. Wians

ROC curves in clinical chemistry: uses, misuses, and possible solutions

Clinical Chemistry, 50 (7) (2004), pp. 1118-1125

View in Scopus ↗ Google Scholar ↗

Ozdamar, 2004 K. Ozdamar

Paket programlarla istatistiksel veri analizi—1

Kaan Kitabevi, Eskisehir (2004)

Google Scholar ↗

Principe et al., 1999 J. Principe, N.R. Euliano, W.C. Lefebvre

Neural and adaptive systems: Fundamentals through simulations

John Wiley & Sons, New York (1999)

Google Scholar ↗

Rahman et al., 2004 M.A. Rahman, H.S. Chaudhury, M.A. Malek, M.A. Khaled

Risk factors of angiographically defined coronary artery disease in Bangladesh

Cardiovascular Pathology, 13 (2004), pp. S80-S138

Google Scholar ↗

Sharma, 1996 S. Sharma

Applied multivariate techniques

John Wiley & Sons, New York (1996)

Google Scholar ↗

Stark and Pfeiffer, 1999 K.D.C. Stark, D.U. Pfeiffer

The application of non-parametric techniques to solve classification problems in complex data sets in veterinary epidemiology—an example

Intelligent Data Analysis, 3 (1999), pp. 23-35

View PDF View article View in Scopus ↗ Google Scholar ↗

The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 2003 The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus

Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus

Diabetes Care, 26 (2003), pp. S5-S20

Google Scholar ↗

Ture et al., 2005 M. Ture, I. Kurt, A.T. Kurum, K. Ozdamar

Comparing classification techniques for predicting essential hypertension

Expert Systems with Applications, 29 (3) (2005), pp. 583-588

View PDF View article View in Scopus ↗ Google Scholar ↗

Turkish Heart Report, 2000 Turkish Heart Report (2000). Turkish society of cardiology. Istanbul: Yenilik Publication, pp. 11–25.

Google Scholar ↗

Zeiher et al., 1993 A.M. Zeiher, H. Drexler, B. Saurbier, H. Just

Endothelium-mediated coronary blood-flow modulation in humans—effects of age, atherosclerosis, hypercholesterolemia, and hypertension

Journal of Clinical Investigation, 92 (2) (1993), pp. 652-662

View in Scopus ↗ Google Scholar ↗

Cited by (333)

Performance analysis of the water quality index model for predicting water state using machine learning techniques

2023, Process Safety and Environmental Protection

Show abstract

An efficient convolutional neural network for coronary heart disease prediction

2020, Expert Systems with Applications

Show abstract

Machine learning-based coronary artery disease diagnosis: A comprehensive review

2019, Computers in Biology and Medicine

Show abstract

Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests ↗

2011, BMC Research Notes

Show abstract

Unobtrusive sensing and wearable devices for health informatics ↗

2014, IEEE Transactions on Biomedical Engineering

Mining electronic health records: Towards better research applications and clinical care ↗

2012, Nature Reviews Genetics

View all citing articles on Scopus ↗

View Abstract