Designing and evaluation of a decision support system for prediction of coronary artery disease

Asieh Khosravanian, MSc [1] Saeed Ayat, PhD [2]

MSc of Department of Computer and Engineering and IT [1], Payame Noor University, Iran. Associate Professor Department of Computer and Engineering and IT [2], Payame Noor University, Iran.

(Received 13 Feb, 2014 Accepted 20 Oct, 2014)

ABSTRACT

Introduction: Since human health is the issue of Medical Research, correct prediction of results is of a high importance. This study applies probabilistic neural network (PNN) for predicting coronary artery disease (CAD), because the PNN is stronger than other methods.

Methods: In this descriptive-analytic study, The PNN method was implemented on 150 patients admitted to the Mazandaran Heart Center, sari. For designing the network, 80% of the data were used for stage of network training, and the remained 20% were used for stage of network testing. In order to implement the network, facilities and functions existing in MATLAB 7.12.0 were used and simulation was conducted in a PC with configurations of corei5 CPU, 2GHz processor, 4GB ram, under operating system of Windows 7.

Results: After 5 times simulation and comparison of the models produced, sensitivity and specificity rates obtained were 1 and 1. In the end, model correctly categorized some healthy subjects who did not need angiography and the treatment related to coronary artery disease.

Conclusion: Due to the high specificity index, this model prevents side effects of angiography in patients who don't need such treatments. Moreover, due to high sensitivity, it can diagnose the patients who really need such diagnostic measures.

Correspondence:
A. Khosravanian, MSc.
Department of Computer and Engineering and IT, Payam Noor University.
Iran
Tel:+98 9171038258
Email:
khosravanian.a@gmail.com

Key words: Coronary Artery Disease, Prediction, Probabilistic Neural Network

Introduction:

The most common heart disease is the coronary artery disease that according to report of the American Heart Association is considered in the list of top ten factors of death in the world (1). The best method for diagnosing this disease is to use angiography (2). However, in this paper, probabilistic neural network has been used to diagnose this disease.

Decision support systems which have helped the human being with solving problems and complicated decision-makings, have recently been highly noticed by many of the researchers for diagnosing diseases. These systems by using data analysis techniques can discover patterns in medical data and improve decision-making process and consequently influence the costs and increase quality of health care (3). There are various techniques for data analysis that as the most common techniques we can mention decision-making tree, Bayes classifier, neural networks, support vector machine, association regulations,

law-based classification, k-nearest neighbor algorithm, rough sets, clustering algorithms, and genetic algorithm (4).

Investigation and study of various references indicate that artificial neural networks have been widely applied in diagnosis of diseases. Of the advantages of using artificial neural networks is that factors such as fatigue, weariness, and emotional situations have no effect on them.

According to reference (5), in some of the studies, results of using modern methods such as neural networks have been observed on classic methods such as regression.

Probabilistic neural networks classify patterns based on Bayesian strategy and non-parametric estimators of probability density functions and are of those kinds of methods that act based on statistical patterns existing in the data. This type of networks use single-pass learning algorithms of data, meaning that the network is only trained in a single course and there is no need for frequent training courses. This is considered as one of the advantages of probabilistic neural network, since the methods which use mental approaches for achieving patterns existing in the data, usually require establishment of a large number of small changes in network parameters to gradually correct its range, and it means that learning term is very long (6).

When an input vector is applied on the network, the first layer calculates distance between input vector from learning inputs, and hence, provides a vector order which are determining elements of the distance extent between input and learning input. The second layer, using output of the first layer, produces a vector of probabilities as output of the network. Finally, the competitive transfer function existing in the second layer, selects the maximum probability extent from vector of probabilities and corresponding to that probability and other probabilities, produces output 1 and 0, respectively.

As mentioned in (5), artificial neural network has been used in solving problems related to clinical diagnosis, analysis of medical pictures, survival prediction, and also has been used in a wide range of medical fields such as oncology, cardiography, hematology, special cares, diagnosis based on medical pictures, infertility, and surgery (6-16).

Particularly, wide research have been conducted in field of coronary artery disease by using artificial neural network that in this part we mentioned a number of them as examples.

In an article with title of "Application of artificial neural network in determination of important predictors of intra-hospital death after coronary artery bypass surgery and comparing it with logistic regression model", an artificial neural network with 19 neurons in the input layer, 4 neurons in the latent layer, and 2 neurons in the output layer by using back propagation error algorithm and Sigmoid transfer function was used (17).

In another article, artificial neural network has been compared with other statistical methods such as logistic regression for predicting coronary artery disease that in this comparison, ROC (Receiver Operating Characteristic) curve was used.

In another study, artificial neural network has been used for diagnosis of difference between artery disease and obstructive pulmonary disease which have similar syndromes among 266 patients in one of the hospitals in Tehran. And multilayer perceptron artificial neural networks and radian basis networks have been used for diagnosis (18).

In the present paper, probabilistic artificial neural network has been used to classify people into two classes of healthy and patient. Implementation of this paper is important since it prevents from probable side effects and damages of angiography and treatments related to coronary arteries of heart for patients who don't need it. Also costs of diagnosis test for such people is reduced.

Methods:

In this descriptive-analytic study which predicts coronary artery situation of heart based on input variables. Statistical population includes people who have undergone angiography treatment in 2010 in heart center of Mazandaran. Results obtained from surveying of the population with 150 samples which have been randomly selected and the questionnaire including demographic information, disease information, diagnosis tests, and angiography result existing in the file completed for them were used.

For implementing this network, MATLAB 7.12.0 software was used and by using cross

validation method the data were divided into five classes and in each stage, 80% of the data were used for network training stage, and the remained 20% were used for network test stage. 14 clinical variables have been used as network input that for this reason a matrix with 14 rows (clinical variables) and 120 columns, for training stage, and another matrix with 2 rows (normal status=0, patient=1) and 120 columns have been used as the target matrix in the training stage.

Input variables are normalized and then applied as input on the probabilistic neural network. In normalization method, variable numbers shall be placed between 0 and 1 so that can be applied to probabilistic neural network. This method of data normalization and classifying them in two different classes has increased accuracy of this network compared to other networks implemented in this regard and was followed by reduction in learning time in the learning period of neural network. Therefore, any target vector has two elements that one of them is zero and the other is one. When input vector applies to the network, its distance from all learning vectors is determined and it is determined that the input is nearer to which one of the learning vectors.

It should be noted that input vector can be applied to the network directly and without need for normalization, and by using existing functions, the matrix can be converted into a matrix consisted of only 1 and 0. Implementation of this method for the neural network designed in this paper indicated that the network has error either in test stage or in learning stage, hence this method was not used.

In this paper, for conducting primary comparison between the two groups that their angiography result was reported as normal and the group that at least one of their vessels was blocked more than 50%, t-test and chi-square test were used similar to reference (5).

Results:

Table 1 indicates characteristics of descriptive statistics related to quantitative and qualitative variables for three states including "all subjects", "normal subjects", and "patient subjects".

Provision of Samples

At first, based on previous studies and expert opinion of cardiovascular specialist, the variables used were determined and from the results obtained from statistics of the 150-subject sample of these variables were used. For establishing a general view of variables, descriptive-analytic statistics related to qualitative quantitative variables were used. Table 1 includes mean value and standard deviation of quantitative variables such as age, creatinine, and ejection fraction which have had significant difference between the two healthy and patient groups, and body mass index, cholesterol, and tri-glyceride didn't have much difference. The reason of lack of difference of cholesterol in the two groups was due to lack of existence of HDL (High Density Lipoprotein) and LDL (Low Density Lipoprotein).

Information related to qualitative variables indicate that except cigarette and disease records, other variables such as sex, high blood pressure, diabetes, heart attach experience, exercise test result, and echocardiography result have had significant relationship with angiography. On the whole, 44 subjects had normal angiography result and 106 subjects had abnormal angiography results (36 subjects had one stenotic artery, 26 subjects had two stenotic arteries, and 44 subjects had three stenotic arteries).

Stage of Training Neural Networks

In this stage, 80% of the data (120 subjects) were used for training artificial neural network. In its selection method, the data were divided into five classes each with 30 subjects for testing the network and each time one of these classes was used for testing and the rest of them were used in training stage. For implementing probabilistic neural network in MATLAB software, an input matrix with 14 rows (variables of table 1) and 120 columns, and another matrix with 2 rows (subject with normal status=0, subject suffering from coronary artery disease=1) and 120 columns was used as target matrix.

Due to using probabilistic neural networks, only one course of training was used for the network, that as mentioned before, this is considered as the main advantages of probabilistic neural networks compared to other networks.

Table 1. Descriptive-Statistical characteristics related to quantitative and qualitative variables for three states including "all subjects", "normal subjects", and "patient subjects"

Variable name		Mean (standard deviation) of "All subjects"	Mean (standard deviation) of "Normal subjects	Mean (standard deviation) of "CAD Patients	Significance Level
Age		57.79 (10.522)	55.6 (10.3)	60.1 (10.3)	0.016
Body Mass Index		26.36 (4.1)	26.9 (3.9)	26.1 (10.3)	0.254
Creatinine		0.91 (0.24)	0.84 (0.19)	0.94 (0.25)	0.016
Cholesterol		193.27 (48.67)	19.04 (32.2)	194.5 (53.8)	0.644
Tri-Glyceride		202.21 (104.05)	208.6 (107)	199.5 (102)	0.628
Eject fraction		44.53 (9.98)	51.13 (7.9)	41.79 (9.5)	0.001
		Total frequency N=150	**Freqency (%) Normal N=44**	**Frequency (%) CAD Disease**	**Significance Level**
Sex	**Male**	67	10 (22.7)	57 (53.8)	0.001
	Female	83	34 (77.3)	49 (46.2)	
Smoking cigarette		29	3 (6.8)	26 (24.5)	0.006
High blood pressure		76	23 (29.5)	53 (59.4)	0.001
Diabetes		50	7 (15.9)	43 (45.8)	0.004
Heart disease family background		38	8 (18.2)	30 (28.3)	0.194
Heart attack background		21	1 (2.3)	20 (18.5)	0.008
Exercise test result		117	15 (34.1)	102 (96.2)	0.001
Echo result		76	14 (31.8)	62 (58.5)	0.003

Table 2. Results of simulation on data set

Data	Performance indexes		Prediction of the system designed		Results of angiography	
	Property	**Sensitivity**	**Normal**	**Coronary artery disease**	**Normal**	**Coronary artery disease**
1st Class	1	1	11	19	11	19
2nd Class	0.909	1	10	20	11	19
3rd Class	1	0.842	14	16	11	19
4th Class	1	0.894	13	17	11	19
5th Class	1	1	11	19	11	19

Figure 2, indicates probabilistic neural network implemented in MATLAB software.

Test Stage of Neural Network

In this stage, 30 subjects which hadn't been used in the training stage, were applied as vector in the probabilistic neural network implemented in the software and this test was repeated for five times. Table 2 indicates results of this simulation in the test stage of the network.

As indicated in the table, of the 30 subjects used in the test stage, 11 subjects had normal angiography results and 19 subjects suffered from the disease. Artificial neural network, also in its best simulation (data of the first and fifth classes), diagnosed 11 subjects with normal result and 19 subjects as patient and this shows that the network has accurately classified healthy and patient subjects. For data of the second class, the artificial neural network diagnosed 10 subjects as normal and 20 subject as patient. Meaning that the designed network has incorrectly diagnosed 4 healthy subjects as patients and also incorrectly diagnosed three patient subjects and healthy ones. In the data of the third and fourth classes, all healthy subjects were correctly diagnosed. However, in the third class data, 3 patients, and in the fourth class data, 2 patients, were incorrectly diagnosed as healthy

subjects. Generally, for investigating efficiency and success extent of classification and disease diagnosis systems, confusion matrix is used. Analyses of confusion matrix in classification and diagnosis of patients results in four states including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). From results of confusion matrix, two characteristics of specificity and sensitivity are obtained which are used for analysis of performance of classification systems.

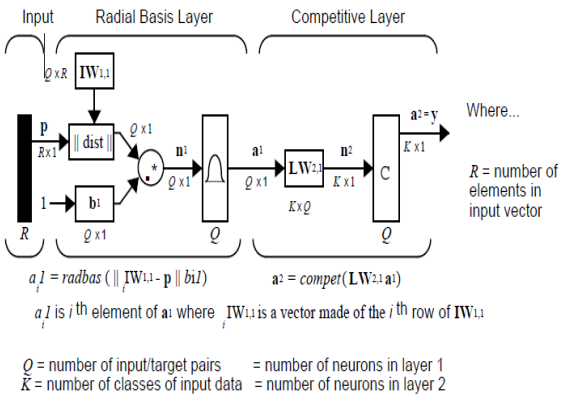

Figure 1. Probabilistic neural network structure

For set of the data used in this paper, sensitivity and specificity are given in table 3, so that in the best situation, these variables are equal to 1 and 1.

Diagonal elements in this matrix indicate number of cases classified correctly. Non-diagonal elements are also elements that are not correctly classified.

Implementation results are shown in table 3 by using tool box of neural network existing in the software for data of the first and fifth classes, and reaction of the probabilistic neural network implemented, has been analyzed by using confusion matrix.

Figure 3 indicates dispersion matrix and various errors about implemented network in the best situation.

80% of data indicated in this matrix is for training purpose which includes 119 subjects, 5% of data is used for validation which includes 8 subjects, 10% of the data is for testing stage which includes 23 subjects, and then the general matrix is indicated for 150 subjects.

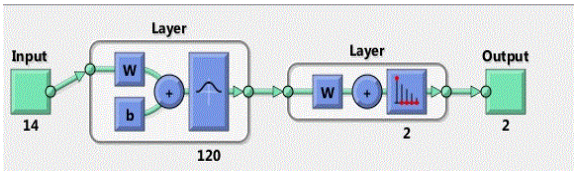

Figure 2. Probabilistic neural network implemented in MATLAB software

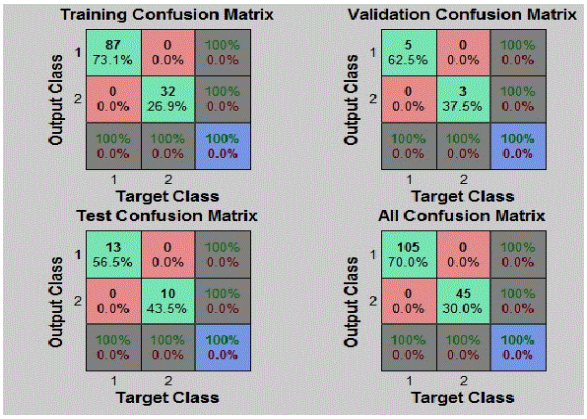

Figure 3- Confusion matrix related to 150 subjects

The ROC curve drawn for the best situation of this implementation, indicates sensitivity extent against specificity in each of the stages.

Figure 4 indicates ROC curve for 150 subjects, using probabilistic neural network (Figure 4).

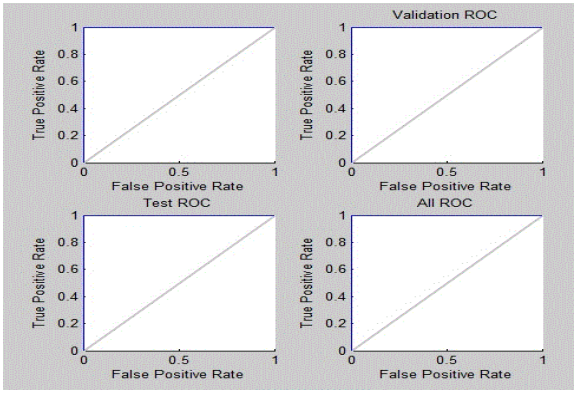

Figure 4. ROC curve for 150 subjects, using probabilistic neural network

Conclusion:

Probabilistic neural network used in this paper can predict patients who don't need angiography and treatments related to coronary artery disease with an appropriate accuracy. According to reference (5), the obtained result is consistent with other research in the world (12).

Mr. Brat Mabli et al., published a paper with title of prediction of coronary artery stenosis by artificial neural network, which has the most consistency with our object of research (14). Data of this research registered from angiography of the patients include 14 input variables (age, sex, race, smoking cigarette, blood pressure, diabetes, body mass index) and some other risk factors, and the output variable was coronary artery stenosis or non-stenosis. Results of this research which had used multilayer perceptron artificial neural network, in the best situation, had evaluated sensitivity as equal to 1, and specificity as equal to 0.47.

Also a research conducted in Iran (5), with title of "application of neural network for evaluating coronary artery disease" which had used MLP artificial neural network with back-propagation algorithm with structure of NN (1), after 1500 training courses, reported sensitivity and specificity as equal to 0.96 and 1, respectively.

Implementation of this paper in using probabilistic neural network has indicated that this network can be created during one course, which is a very short period compared to 1500 training courses in a similar case mentioned in reference (5).

In fact, implemented network in this paper, due to high speed and good extendibility is superior to back-propagation neural networks. Training required for this network is much different as faster than training algorithm required for artificial neural network with back-propagation algorithm. In this network, training process includes one stage and there is no iteration needed for adjustment of weights which is against back-propagation method which needs great number of iterations for adjustment of weights.

After five times of simulation, results indicate that the network is designed in a manner that can indicate appropriate reaction to non-training data, in other words, the network has an appropriate generality.

In this research, sensitivity and specificity parameters in the best situation by help of this network, were equal to 1 and 1, and it indicates that this network classifies the data with appropriate accuracy and can be an appropriate substitute for angiography. Of the reasons of high values of sensitivity and specificity in the present paper, we can mentioned normalization of input vector and appropriate selection of neural network for this purpose. This result is of a high importance since it prevents from probable side effects and damages of angiography for patients who don't need it. Also it can determine the patients who really need immediate diagnosis treatments in the shortest time and with the highest accuracy. Furthermore, the implemented software, by using this simulation can be used for training physicians. In the future, some modifications can be applied in the implementation as follows. Since all related factors in mentioned in table 1 have equal importance (weight), they can be weighted so that more important factors have more weight and obtained results be considered.

References:

1. American Heart Association. Available from: URL: http://www.heart.org/HEARTORG.
2. Crawford M. Current diagnosis & treatment in cardiology. 3rd ed. NewYork: McGraw-Hill Press; 2009.
3. Teixeira-Pinto A, Normand SL. Correlated bivariate continuous and binary outcomes: issues and applications. Stat Med. 2009;28(13):1753-1773.
4. Mahmoudi I, Askari moghadam R, Moazzam M, Sadeghian S. Prediction model for coronary artery disease using neural networks and feature selection based on classification and regression tree. Shahrekord University of Medical Sciences. 2013;15(5):47-56.
5. Mohammadpour R, Esmaeili M, Ghaemia A, Esmaeili J. Application of Artificial Neural Network for Assessing Coronary Artery Disease. Mazandaran University of Medical Sciences. 2012;21(86):7-9. [Persian]

6. Menhaj MB. Computational intelligence, fundamentals of neural network. 2nd ed. Tehran: Amir Kabir University Press; 2008. [Persian]

7. George J, Ahmed A, Patnaik M, Adely Y, Levy Y, Harats D, et al. The prediction of coronary atherosclerosis employing artificial neural networks. Clin Cardiol. 2009;23(6):453-456.

8. 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):366-374.

9. Falk CT. Risk factors for coronary artery disease and the use of neural networks to predict the presence or absence of high blood pressure. BMC Genet. 2003;4(Suppl 1):S67.

10. Itchhaporia D, Snow PB, Almassy RJ, Oetgen WJ. Artificial neural networks: Current status in cardiovascular medicine. J Am Coll Cardiol. 1996;28(2):515-521.

11. Chong CF, Li YC, Wang TL, Chang H. Stratification of adverse outcomes by preoperative risk factors in coronary artery bypass graft patients: an artificial neural network prediction model. AMIA Annu Symp Proc. 2003;160-164.

12. Babaoglu I, Baykan OK, Aygul N, Ozdemir K, Bayrak M. Assessment of exercise stress testing with artificial neural network in determining coronary artery disease and predicting lesion localization. Expert Syst Appl. 2010;37(1):905.

13. Ohno-Machado L, Musen MA. Sequential versus standard neural networks for pattern recognition: An example using the domain of coronary heart disease. Comput Biol Med. 1997;27(4):267-281.

14. Mobley BA, Schechter E, Moore WE, McKee PA, Eichner JE. Predictions of coronary artery stenosis by artificial neural network. Artif Intell Med. 2000;18(3):187-203.

15. Dirgenali F, Kara S. Recognition of early phase of atherosclerosis using principles component analysis and artificial neural networks from carotid artery Doppler signals.Expert Syst Appl. 2006;31(3):643-651.

16. Harrison RF, Kennedy RL. Artificial Neural Network Models for Prediction of Acute Coronary Syndromes Using Clinical Data From the Time of Presentation. Ann Emerg Med. 2005;46(5):431-439.

17. Biglarian A, Babaee GR, Azmie R. Application of Artificial Neural Network Model in Determining Important Predictors of In-Hospital Mortality After Coronary Artery Bypass Graft Surgery, and it's Comparison with Logistic Regression Model. Modarres J Med Sci. 2004;7(1):23-30.

18. Mehrabi S, Maghsoudloo M, Arabalibeik H. Application of multilayer perceptron and radial basis function neural networks in differentiating between chronic obstructive pulmonary and congestive heart failure diseases. Expert Systems with Applications. 2009;36(3):6956-6959.

طراحی و ارزیابی یک سیستم پشتیبان تصمیم برای پیش‌بینی بیماری عروق کرنری قلب

آسیه خسروانیان [1] دکتر سیدسعید آیت [2]

[1] کارشناس ارشد گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور [2] دانشیار گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور

مجله پزشکی هرمزگان سال نوزدهم شماره ششم بهمن و اسفند ۹۴ صفحات ۴۰۸-۴۰۱

چکیده

مقدمه: *ازآنجا که در تحقیقات علوم پزشکی مسئله سلامت انسان مطرح است، پیش‌بینی درست نتایج اهمیت بیشتری می‌یابد. در این مقاله از شبکه عصبی احتمالی جهت پیش‌بینی وضعیت عروق کرنری قلب استفاده شده است.*

روش کار: *در این مطالعه توصیفی - تحلیلی، جامعه آماری شامل ۱۵۰ نفر بیمار مرکز فوق تخصصی قلب مازندران بود. مدل پیش‌بینی وضعیت عروق کرنری قلب با استفاده از شبکه عصبی احتمالی (PNN) تولید شد. برای طراحی شبکه، از ۸۰ درصد داده‌ها جهت مرحله آموزش شبکه و ۲۰ درصد باقیمانده جهت مرحله آزمون شبکه استفاده شده است. به منظور پیاده‌سازی شبکه از امکانات و توابع موجود در نرم‌افزار متلب نسخه ۷/۱۲/۰ بهره گرفته شده و بر سیستم corei5 با پردازنده 2.4 GHz و حافظه 4GB تحت ویندوز ۷ شبیه‌سازی انجام شده است.*

نتایج: *پس از ۵ مرتبه شبیه‌سازی و مقایسه مدل‌های تولیدشده، عملکرد شبکه عصبی احتمالی پیاده‌سازی شده بر اساس شاخص‌های عملکردی اختصاصیت (specificity) و حساسیت (sensitivity)، در مرحله آزمون شبکه معادل عدد یک به دست آمد و در نهایت توانستیم افراد سالم و افرادی که دچار بیماری عروق کرنری بودند را با دقت بهتری نسبت به موارد مشابه قبلی طبقه‌بندی کنیم.*

نتیجه‌گیری: *اختصاصیت و حساسیت به‌دست‌آمده از این شبیه‌سازی نشان داد که استفاده از شبکه عصبی احتمالی می‌تواند جایگزین مناسبی برای آنژیوگرافی در پیش‌بینی بیماری عروق کرنری قلب باشد و از عوارض و آسیب‌های احتمالی آنژیوگرافی جلوگیری نماید.*

کلیدواژه‌ها: *بیماری عروق کرنری، پیش‌بینی، شبکه عصبی احتمالی*

نویسنده مسئول:
آسیه خسروانیان
گروه مهندسی کامپیوتر و فناوری اطلاعات،دانشگاه پیام‌نور
ایران
تلفن: ۰۹۱۷ ۱۰۳۸۲۵۸ ۹۸+
پست الکترونیکی:
khosravanian.a@gmail.com

دریافت مقاله: ۹۲/۱۱/۲۴ اصلاح نهایی: ۹۳/۷/۶ پذیرش مقاله: ۹۳/۷/۲۸