Evaluation of Risk Factors Selection in Cardiac Risk Stratification

E Yargholy[1,2], S Parvaneh[1,2]

[1]Islamic Azad University, Science and Research Branch, Tehran, Iran
[2]Young Researchers Club, Tehran, Iran

Abstract

In the past few years a number of algorithms for cardiovascular risk stratification have been proposed to the medical community but a big question has been remained unsolved: From among alternative sets of cardiac risk factors which ones are more significant in cardiac risk stratification? In this paper a hybrid neuro-fuzzy classifier, IRIDIA Method for Neuro-fuzzy Identification and Data Analysis, was applied to determine the cardiac risk stratification. Using the obtained neuro-fuzzy model, sensitivity analysis of cardiac risk factors was derived to sort out cardiac risk factors according their significance - which factor has great effect on cardiac risk.

1. Introduction

The concept of cardiac risk factors has become an integral part of the modern medical curriculum and has led to the development of algorithms for cardiac risk assessment [1]-[4].

Usually, the cardiac risk stratification methods in literature are either based on ECG data-driven or patient's medical data. The former includes QRS detection, distances in time between waves of RR, PR, and etc [3], [4]. Age, blood pressure, cholesterol, and … are examples of the latter one [1]-[2].

In the previously done researches, data have been measured, then several score sheets have been prepared accordingly [5], [6]. These score sheets can be applied to estimate people risk factor in general. The patient's data are scored by means of obtained score sheets. Considering the scores, cardiac risk factor is calculated.

Framingham study is regarded as the earliest method in this era. The objective of Framingham heart study was to identify the common factors or characteristics that contribute to cardiovascular disease by following its development in the large group of participants who had not yet developed overt symptoms of cardiovascular disease or suffer a heart attack. The researchers went through medical history, extensive physical examination, and laboratory tests.

Over the years, careful monitoring of Framingham study population has led to the identification of major cardiovascular disease risk factors, as well as valuable information on the effects of these factors such as blood pressure, blood triglyceride and cholesterol levels, age, and gender [7].

The use of scoring systems to stratify cardiac risk is currently hampered by methodological and mathematical flaws. Newer approaches make use of fuzzy logic and artificial neural networks, linked to artificial intelligence. Applying neuro-fuzzy method, the present study seems to initiate a new approach toward cardiac risk stratification.

So far, a lot of numeric analysis approaches of neuro-fuzzy systems have been presented. Most of these techniques divide the input data space, not considering the output data space, so the obtained rules should not be always reasonable. IRIDIA method is a simple and effective neuro-fuzzy network that divides input-output data space and extracts fuzzy if-then rules automatically. This method searches for the right number of rules of the architecture by adopting a procedure of cross-validation on the available data set [8], [9].

To aim at excellent fuzzy modeling, it is necessary to select adequate input elements. Unnecessary inputs interfere with proper learning, inference, and creation of efficient rules [10]. Sensitivity analysis of cardiac risk factors was derived by using the obtained fuzzy rules in order to show which factor has great effect on cardiac risk. When the factors are uncertain, and/or there are alternative sets of factors to choose from, the model output will also be uncertain. Quantitative sensitivity analysis tries not only to identify but also to quantify the relative significance of the factors [11]. Selection of efficient input elements can contribute to the performance improvement, reduction of calculation cost and analysis of the obtained rules that is one of the important merits of fuzzy system.

2. Database

The present article develops a simplified cardiac risk factor model, building on the sex, age, cholesterol (LDL), and blood pressure. Family history for heart disease, physical activity, obesity, cigarette smoking, diabetes and … are not included because these factors work to a large extent through a major risk factors, and their unique contribution to cardiac risk factor assessment can be difficult to quantify [12]-[14].

Statistic data of 165 patients in hospitals of Tehran (54 patients) and Shiraz (111 patients)[1] were gathered. The data included the following variables: sex, age, LDL, blood pressure, and Myocard-brain Creatinine Phosphokinase (cpk-mb) enzyme. The intensity of infarction was determined according to the amount of the enzyme (in three different levels: no infarction, mild infarction, and severe infarction).

3. Idea and method

Sex, age, LDL, and blood pressure were considered as the input and infarction intensity was considered as the output.

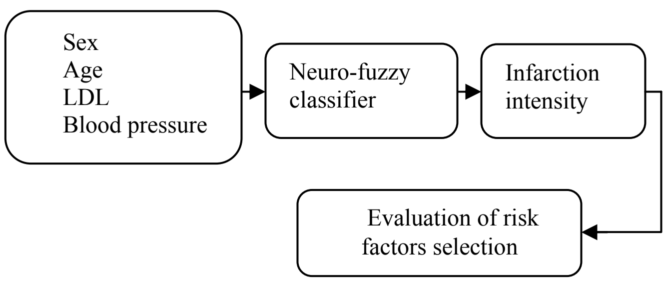

Fig. 1. Assessment system

To draw the input-output mapping of each group, neuro-fuzzy networks were used. Having obtained a neuro-fuzzy model, we estimated the significance of input variables in making output with the use of Linkens method (Fig. 1.).

3.1. Cardiac risk stratification

At first step, in this study all the input variables including age, LDL blood pressure were normalized in the range of [0,1]. Output values of no infarction, mild infarction, and severe infarction based on enzyme concentration are identified with numerical values of 0, 0.5 and 1 respectively. Then the data related to men, women, men-Shiraz, women-Shiraz, men-Tehran, and women-Tehran were partitioned into train and test sets randomly (both sets included subjects with no infarction, subjects with mild infarction as well as subjects with severe infarction). %70 of whole data was used for training (rule generation) and the rest %30 was used for testing. IRIDA method was applied to data sets of men, women, men-Shiraz, women-Shiraz, men-Tehran, and women-Tehran.

IRIDIA Method. The precise model of Takagi-Sugeno selected to be used as the fuzzy model. The program provided a set of structural alternatives in the definition of the fuzzy model: the shape of membership function of the antecedents (Gaussian/triangular), the parametric form of the consequent (constant/linear), the combination method of the rules (weighted/comb) and bias term which can be either used or not in the linear step of the nonlinear parametric optimization (bias/no-bias). Further, it is possible to choose between two different clustering policies, K-means/Hyperplane fuzzy clustering [15], to provide the identification algorithm with good initial values of the center and base of the membership functions. Then we obtained 32 models for each data set:

(Gaussian/Triangular,Constant/Linear,Weighted/Comb , Bias/No-bias,K-means/Hyperplane).

IRIDIA method searches for the right number of rules of the architecture by adopting a procedure of cross-validation on the available data set. It starts with a minimal number of rules and at each step increases the number of rules by restarting the global procedure until a maximum number of rules is reached (the user is free to set properly what is the desired range of number of rules to range over). Then each structure is characterized by its error in generalization, estimated by a procedure of cross-validation and the optimal number of rules is searched by comparing the cross-validation error obtained at different levels of number of rules. At the end of the global training phase, the cross-validation error is plotted against the number of rules used and the user is asked to choose the number of rules at which the fuzzy system looks to perform better (Fig. 2.).

In order to have a definite range for the number of rules, points with coordinates of (blood pressure, LDL, age), were exhibited in a 3-dimentional space. Therefore regarding the distribution of data a range of 3-10 was chosen. For each of 32 fuzzy structures cross-validation error versus number of rules was plotted and number of those rules was chosen that produces the least cross-validation error.

Selecting the fuzzy model, the number of rules was determined. The model was trained by the train set data

[1] . Tehran and Shiraz are two big cities in Iran which are far away from each other.

The trend was repeated for the 32 fuzzy structures.

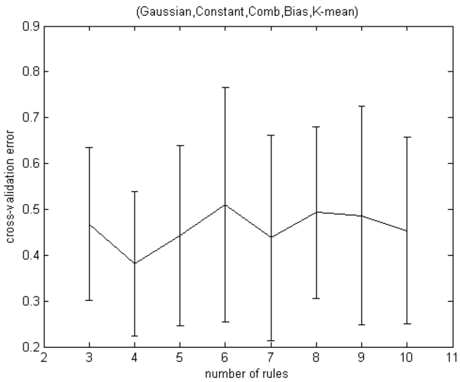

Fig. 2. Cross-validation error versus number of rules diagram

In these phase, we used the trained fuzzy model to predict data output of train and test sets data. Considering the number of errors, we chose the most appropriate model.

Applying IRIDIA method led us to consider a distinction between males and females in cardiac risk stratification. Males and females were studied separately because they differ according to the age at which there is risk of infarction, but in the case of blood pressure and LDL, the differences are not considerable.

The results of the study directed us to regard subjects of different cities separately. Since life quality and conditions such as air pollution, life style and … differ in these two cities, so risk ranges of different age, LDL, and blood pressure levels are not the same in different cities and various adaptive strategies can be tuned up in terms of a specific population to stratify the risk of cardiac disease.

Table 1
Fuzzy models of IRIDIA method

Group	Fuzzy model	Number of rules	Total error
Men	Gaussian, Constant, Comb, Bias, K-mean	5	%10.31
Women	Gaussian, Constant, Comb, Bias, HFC	5	%14.71
Men-Shiraz	Gaussian, Constant, Weighted, No-bias, K-mean	5	%4.76
Women-Shiraz	Gaussian, Constant, Weighted, No-bias, HFC	7	%2.08
Men-Tehran	Gaussian, Constant, Comb, Bias, K-mean	4	%14.71

Results of different groups of our research are summarized in Table 1. Total error is the proportion of error numbers to the total number of data.

Tehran women data, in our database, was not enough to be applied to IRIDIA method.

3.2. Evaluation of risk factor significance

In this research, we used Linkens [16], [17] method to compute input significance. Since this method has to know the set of proper rules in advance, we first chose the neuro-fuzzy model and then computed the input significance.

Suppose that the dimension of the data, is N_1, the number of rules, is N_2, and the dimension of the output data, is N_3. The input data is expressed as follows:

$U= (u_1, u_2, \ldots, u_{N_1})$.

The subscripts i, j, and k refer to the input, the rule, and the output respectively. u_{ij} is the membership value and w_{ij} is the center value of the membership function. The logical form of the fuzzy inference if-then rules is given such as : if u_1 is w_{1j}, and …, u_i is w_{ij}, … , u_{N_1} is $w_{N_1 j}$ is then y_k is w_{jk}.Where w_{ij} means the value near w_{ij}. The lth input significance value is calculated.

$$z_{ik}^{l} = \frac{\sum_j^{N_2} u_{ij} w_{ik}}{\sum_j^{N_2} u_{ij}} \tag{1}$$

Next, the integrated input significance value from the ith input to the kth output is calculated using L sets of z_{ik}^{l}.

$$R_{ik} = \max_l (z_{ik}^{l}) - \min_l (z_{ik}^{l}) \tag{2}$$

$$R_k = \max_i (R_{ik}) \tag{3}$$

Finally, the syntactic input significance value is calculated by accumulating F_{ik}.

$$F_i = \sum_k^{N_3} F_{ik} \tag{4}$$

4. Results

With regards to table 2, it can be concluded that cardiac risk factor significance differs in terms of gender and regional conditions.

The significance of input variables (blood pressure, LDL, and age) was approximately the same. So our input set has been appropriate and there is no need to decrease the number of variables or change them. Age, blood pressure and LDL, input variables, were sorted out according their significance respectively. With the benefit of this study, the significance degree of a risk factor can be defined according to a specific population and specific region.

Table 2
Significance of input variables

Group	age	L.D.L.	Blood Pressure
Men	1	0.9997	0.9997
Women	0.9958	1	1
Men-Shiraz	1	0.9946	0.9985
Women-Shiraz	1	0.9873	0.9986
Men-Tehran	1	1	1

For further research, additional risk factors, not available in patients' records at the present, can be included to provide the medical community with a comprehensive program which shows the risk factors significance in a hierarchical order.

5. Conclusion

In this paper we explained cardiac risk factor stratification based on a neuro-fuzzy classifier, IRIDIA method. Our Parameters consisted of sex, age, LDL, and blood pressure. Sensitivity analysis was used to determine: the resemblance of this model with the process under study, the quality of model definition, factors that mostly contribute to the output variability and interaction between factors.

The capabilities of our method can be summarized as below:

- More Risk factor, smoking background, life condition, obesity, diabetes, and … can be applied in the proposed method.
- Extension of available data in database simply can be integrated in our model. Also, Extension of database with more data can yield more reliable results.
- Using this method, risk factor assessment can be specialized based on different regions and life conditions.
- Defining the most effective factor, in each special population, prevention strategies can be adopted.
- Risk factor can be compared in different populations and differences can be derived.

References

[1] Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow up of the prospective cardiovascular unster(PROCAM)study. Circulation. J. 2002;105:310-5.

[2] Menotti A, Puddu PE, Lant M. Comparison of the Framingham risk function-based coronary chart with risk function from an Italian population study. Eur Heart. J. 2000;21:365-70.

[3] Kennedy R *et al*. Early diagnosis of acute myocardial infarction using clinical and electrocardiographic data at presentation: Derivation and evaluation of logistic regression models. Eur Heart J.1996 ;17: 1181-91.

[4] Green M *et al.* Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in emergency room. Artificial Intelligence in Medcine 2006;38:305-18.

[5] De Backer G, Ambrosioni E, Borch-Jonhsen K *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 representative of eight societies and invited experts). Atherosclerosis 2004;173:381-91.

[6] National Cholesterol Education Program expert panel. 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): final report. National Institute of Health III. 2002. National Heart and Lung Institute. NIH Publication 02-5215 September 2002.

[7] www.Framinghamheartstudy.org

[8] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and selection. 14thIinternational Joint Conference on Artificial Intelligence 1995:1137-43.

[9] Bontempi G, Birattari M. Matlab software tool for neuro-fuzzy identification and data analysis, ftp://iridia.ulb.ac.be/pub/lazy/paper/IridiaTr 1999-09.ps.gz.

[10] Iyatomi H, Hagiwara M. Adaptive fuzzy inference neural network. Pattern Recognition 2004;37: 2049-57.

[11] Saltelli et al. Global Sensitivity Analysis 2007. New York: The Primer. Wiley.

[12] Gordon T, Kannel WB. Multiple risk function for predicting coronary heart disease. Am Heart. J.1982;103:1031-9.

[13] Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. Diabetes, blood lipids, and the role of obesity in coronary heart disease risk for women. Ann Inten Med 1977; 87: 393-7.

[14] www.intelihealth.com

[15] Bezdek JC, Anderson IM. An application of the c-varieties clustering algorithms to polygonal cure fitting. *IEEE Trans.* Systems, Man Cybern 1985;15: 637–41.

[16] Linkens DA, Chen MY. Input selection and partition validation for fuzzy modeling using neural network. Fuzzy Sets System , 1999;107: 299-308.

[17] Chen MY, Linkens DA. A systematic neuro-fuzzy modeling framework with application to material property prediction. *IEEE Trans.* Systems, Man Cybern 2001;31: 781–90.

Address for correspondence
Elahe Yargholy
Iran, Tehran, Sardare Jangal Ave., Golzare1 Ave.
Islamic Azad University, Science and research Branch
School of Biomedical Engineering
E-mail address: elahe_yargholy@yahoo.com