Novel Cardiac Risk Factor Stratification Using Neuro-fuzzy Tool

Elahe Yargholi
*Islamic Azad University, Science and
Research Branch, Tehran, Iran
Young Researchers Club, Tehran, Iran*
elahe.yargholi@gmail.com

Saman Parvaneh
*Islamic Azad University, Science and
Research Branch, Tehran, Iran
Young Researchers Club, Tehran, Iran*
saman.parvaneh@gmail.com

Abstract

Heart disease continues to be leading cause of morbidity and mortality among adults all over the world. Cardiac risk factor assessment requires a classification system that is robust to the interaction and uncertainty of input factors, as well as being interpretable on the decision made. To meet the requirements, we made use of neuro-fuzzy methods that is a certain novelty in cardiac risk assessment. In this paper two hybrid neuro-fuzzy classifiers, Adaptive Network-based Fuzzy Inference System (ANFIS) and IRIDIA method for neuro-fuzzy identification and data analysis, were applied to determine the cardiac risk factor. These two methods are widely used in the area of decision making. The study demonstrated that the IRIDIA method is efficient in cardiac risk factor assessment.

1. Introduction

In the past few years a number of algorithm for cardiovascular risk assessment has been proposed to the medical community [1]-[9].

Cardiac risk factor assessment is the estimation of future cardiac health outcome of people considering several important risk factors. One of the essential tasks is to enable the classification of the outcome (no infarction, mild infarction, or sever infarction) based on known respective outcome, for subsequent patients whom the system has not been previously subjected to.

Usually, the cardiac risk assessment 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, etc [10], [11]. Age, blood pressure, and cholesterol are examples of the latter one [1]-[6].

In the previously done researches, data have been measured, then several score sheets have been prepared accordingly [12], [13]. 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 [14].

Newer approaches make use of fuzzy logic and artificial neural networks, linked to artificial intelligence [15], [16]. In comparison with scoring method, neuro-fuzzy is more likely to lead us toward superior classifying applications. So, on the case of neuro-fuzzy merits, we have initiated to apply it in assessment of cardiac risk factor.

The advantages of fuzzy classifiers have been clearly exhibited in at least two ways: they allow multiple gradual class membership to smooth the transition from one class to the other as the input changes [17], therefore result in a better response in those cases that may overlap over several classes; and they also favor interpretation.

One of the important design problems of a fuzzy system is to instruct a set of appropriate fuzzy rules. Because pure fuzzy system fails to automatically acquire knowledge from data in order to extract the required rules, the researchers had to refer to expert physicians [18] but in present study, the basic idea

1199

behind the design is to estimate fuzzy rules through learning from input-output data pairs. Neural network is introduced to solve this problem [19], [20]. The utility of neural network lies in the fact that they can be used to infer a function from observations and also to use it. This is particularly useful in the applications where the complexity of data or task makes the design of such a function by hand impractical.

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 [21], [22]. Whereas ANFIS, determining the number of membership functions of input variables, uses either, all possible rules or we, should set the rules, ourselves [23]-[25].

2. Materials and methods

2.1. Materials

Extensive clinical and statistical studies have identified several factors that increase the risk of coronary heart disease and heart attack. Major risk factors are those that research has shown significantly increase the risk of heart and blood vessel (cardiovascular) disease. Other factors are associated with increased risk of cardiovascular disease, but their significance and prevalence haven't yet been precisely determined. They're called contributing risk factors. The American Heart Association has identified several risk factors. Some of them can be modified, treated or controlled, and some can't. The more risk factors you have, the greater your chance of developing coronary heart disease. Also, the greater the level of each risk factor, the greater the risk [26].

Increasing age, male sex (gender), and heredity (including race) are the major risk factors that can't be changed. Tobacco smoke, high blood cholesterol, high blood pressure, physical inactivity, obesity and overweight, and diabetes mellitus are the major risk factors you can modify, treat or control by changing your lifestyle or taking medicine. Stress, alcohol, and diet and nutrition are other factors contribute to heart disease risk.

The present article develops a simplified cardiac risk factor model, building on the sex, age, cholesterol (LDL) and blood pressure (systolic). Family history of heart disease, physical activity, obesity, cigarette smoking, diabetes and other probable factors 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 [27]-[29] but they should not be underestimated. For example, individuals with family history of heart disease should view increased risk factors with particular concern and deal with them more vigorously than should people without such backgrounds.

Predicted risk may be calculated with SBP (Systolic Blood Pressure) or DBP (Diastolic Blood Pressure). Because of high correlation between these two measurements, both can not be included; the redundancy leads to difficulty in interpretation. The model incorporating SBP is recommended because SBP is more accurately determined, has a wider range of values, and is a stronger predictor of heart disease, particularly in elderly [30].

Statistic data of 165 patients in hospitals of Tehran (54 patients) and Shiraz (111 patients) [1] were obtained. The patients were chosen from among people who had referred to hospital having of chest pain. They underwent laboratory tests and their blood pressure was measured. The data included the following variables: sex, age, LDL, SBP, 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 sever infarction) [31], [32].
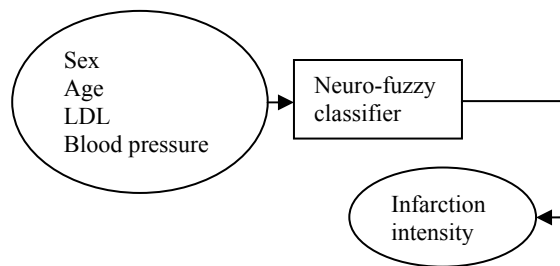

Figure 1. Assessment system

Sex, age, LDL, and SBP were considered as the input and infarction intensity was considered as the

---

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

1200

output. To draw the input-output mapping of each group, neuro-fuzzy networks were used (Fig. 1.).

2.2. Methods

At first step, in this study all the input variables including age, LDL, and SBP were normalized in the range of [0,1]. Output values of no infarction, mild infarction, and sever 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 sever infarction). %70 of whole data was used for training (rule generation) and the rest %30 was used for testing. ANFIS and IRIDA methods were applied to data sets of men, women, men-Shiraz, women-Shiraz, men-Tehran, and women-Tehran.

2.2.1. ANFIS. ANFIS toolbox of MATLAB software was used. We selected Grid partition method for data clustering then all possible rules were used. Fuzzy model of Takagi-Sugeno was applied. For membership functions of antecedent of rules Gaussian/Triangular options were considered and regarding consequent of rules linear/constant were our choices. For error tolerance, two values 0.1 and 0.01 were considered to investigate the possibility of overfitting. We obtained 8 models for each of data sets: (membership function Gaussian/triangular, output function constant/linear, error tolerance of 0.1 / 0.01)

Also we increased the number of Epochs till the error value did not decrease any more.

Having fuzzy models trained, we applied train and test data sets to the model and obtained output of model. The output values less than 0.25 were considered as 0 (people with no infarction), the output values more then 0.25 and less than 0.75 were considered as 0.5 (people with mild infarction), and the output values more than 0.75 were considered as 1 (people with sever infarction). Then we counted the number of cases not correctly predicted by our models. Considering the number of errors, we chose the most appropriate model.

2.2.2. IRIDIA Method. The precise model of Takagi-Sugeno which focuses on accuracy was selected to be used as the fuzzy model [33]. 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 [34], 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 adapting a procedure of cross-validation on the available data set. It stars 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.).
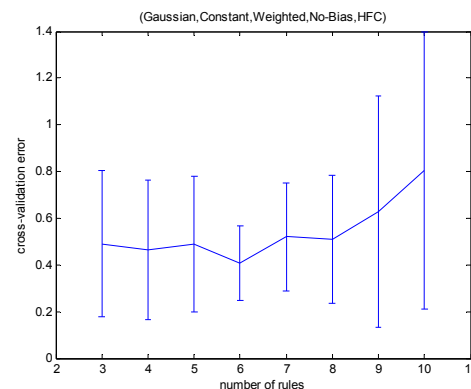

Figure 2. Cross-validation error versus number of rules diagram

In order to have a definite range for the number of rules, points with coordinates of (SBP, 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 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 then center and base of membership functions of rules antecedents and parameters of rules consequents were derived. The trend was repeated for the 32 fuzzy structures.

In these phase, we used the trained fuzzy model to predict data output of train and test sets data similar to ANFIS Method. If the number of errors of two models were the same, the model having fewer rules was chosen to have less complexity in our models.

3.Results

Why male and female were studied separately. Applying neuro-fuzzy method led us to consider a distinction between males and females in cardiac risk stratification. Male and female were studied separately because they differ according to the age in which there is risk of infarction, but in the case of SBP and LDL, the differences are not considerable.

ANFIS and IRIDIA methods were applied to men and women data sets separately to achieve input-output mapping. Then, we considered the men data set as the input of models obtained for women in previous stages and vice versa (men data into women models, women data into men models, men-Shiraz data into women-Shiraz models, women-Shiraz data into men-Shiraz models, men-Tehran data into women-Tehran models). Because of above mentioned reason, the error value increased in almost all cases. Results of stratifying risk factor of one sex using models obtained for the other sex are summarized in Table 1.

Table 1. Stratifying risk factor of one sex using models obtained for the other sex

Group	IRIDIA	ANFIS
Men	%3.09	%27.84
Women	%14.71	%25
Men-Shiraz	%1.6	%-1.59
Women-Shiraz	%8.33	%0
Men-Tehran	-	%23.52
Women-Tehran	-	%60

Why subjects in Tehran and Shiraz were studied separately. ANFIS and IRIDIA methods were applied to men and women data sets separately to derive fuzzy model. With regards to applied methods and previous record of neuro-fuzzy models performance, we expected fewer errors than what we got.

Having studied more we found that the unexpected error values were due to the distribution of data. Our data had been gathered from two distant cities, Tehran and Shiraz. Life quality and conditions such as air pollution, life style and … differ in these two cities, so risk ranges of different age, LDL, and SBP levels are not the same in Shiraz and Tehran. So we repeated the methods for four groups: men-Shiraz (63 people), women-Shiraz (48 people), men-Tehran (34 people), and women-Tehran (20 people).

Results of different groups of our research are summarized in Table 2 and Table 3. Total error is the proportional of error numbers to the total number of data.

Table 2. Fuzzy models of ANFIS

Group	Fuzzy model	Number of rules	Total error
Men	Gaussian, Linear	27	%22.68
Women	Triangular, Linear	27	%19.11
Men-Shiraz	Triangular, Constant	27	%3.17
Women-Shiraz	Triangular, Constant	27	%2.08
Men-Tehran	Gaussian, Linear	27	%37.5
Women-Tehran	Gaussian, Linear	27	%5

Table 3. 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

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

The results show that the IRIDIA method provides us with more accurate information in cardiac risk factor assessment. Gathering data from different regions with different climatic, social, cultural, economic, employment and other circumstances, we can extend this research. Then the appropriate model can be derived in order to estimate cardiac risk factors in different regions.

4. Conclusion

In this paper we explained cardiac risk factor stratification based on neuro-fuzzy method. Our Parameters in this method were sex, age, LDL, and SBP.

In this paper, two hybrid neuro-fuzzy classifiers (ANFIS and IRIDIA method) were applied to determine the cardiac risk factor. The study demonstrates that the IRIDIA method is more efficient than ANFIS in cardiac risk factor data analysis.

The capabilities of our method can be summarize as below:

1- New Risk factor such as smoking background, life condition and … can be included in the proposed method easily.

2- Extension of available data in database simply can be integrated in our model. Also, Extension of database with more data can yields more reliable results.

3- With our method, risk factor assessment can be specialized based on different regions and life conditions.

4- Using the obtained fuzzy rules, sensitivity analysis of cardiac risk factors can be derived to show which factor have great effect on cardiac risk.

5- Comparison of risk factor in different community can be done.

Relation of sex, age, LDL that used in our model with feature extracted in ECG (Clinical symptoms of heart disease) do not evaluated in our work.

The predictions may not be appropriate for individuals with extremely elevated risk factors such as malignant hypertension, extremely high cholesterol levels and … that place them in the top or bottom few percentiles of the distribution.

References

[1] P. W. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, W. B. Kannel, "Prediction of coronary heart disease using risk factor categories", Circulation. J., 97, 1998, pp. 1837-1847.

[2] G. Assmann, P. Cullen, H. Schulte, "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., 105, 2002, pp. 310-315.

[3] A. Menotti, P. E. Puddu, M. Lant, "Comparison of the Framingham risk function-based coronary chart with risk function from an Italian population study", Eur Heart. J., 21, 2000, pp. 365-370.

[4] R. M. Coronary, K. Pyoralak, A. P. Fitzgerald, S. Sans, A. Menotti et al, "SCORE project group estimation of ten-year risk of fatal cardiovascular disease in Europe", Eur Heart. J., 24, 2003, pp. 987-1003.

[5] F. Gonzalez-Velazquez, G. F. Mendez, "Cardiovascular risk stratification by means of the SCORE system in health care workers in Veracruz, Mexico", International Journal of Cardiology, 121, 2007, pp. 81-83.

[6] N. G. Forouhi, N. Sattar, T. Tillin, P. M. McKeigue, N. Chaturvedi, "Do known risk factors explain the higher coronary heart disease mortality in South Asian compared with European men? Prospective follow-up of the Southall and Brent studies, UK", Diabetologia., 49, 2006, pp. 2580-2588.

[7] A. Menotti, P. E. Puddu, M. Lant, "The estimate of cardiovascular risk. Theory, tools and problems", Ann Ital Med Int., 17, 2002, pp. 81-94.

[8] S. Giampaoli, L. Palmieri, P. Chiodini, M. Ferrario, S. Panico et al, "La carta del rischio cardiovascolare globale", Ital Heart. J., 5, 2004, pp. 177-185.

[9] M. T. La Rovere, G. D. Pinna, S. H. Hohnloser, "Baroreflex sensitivity and heart rate variability in the identification of patients at risk for life-threatening arrhythmias," Circulation. J., 103, 2001, pp. 2072-2077.

[10] R. Kennedy 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., 17, 1996, pp. 1181-91.

[11] M. Green et al, "Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in emergency room", Artificial Intelligence in Medcine, 38, 2006, pp. 305-18.

[12] G. De Backer, E. Ambrosioni, K. Borch-Jonhsen 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, 173, 2004, pp. 381-391.

[13] 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.

[14] www.Framinghamheartstudy.org/about/history.html

[15] E. Grossi, "How artificial intelligence tools can be used to assess individual patient risk in cardiovascular disease: problems with the current methods", BMC cardiovascular Disorders, 2006, pp. 6-20.

[16] A. piegat, "Fuzzy modeling and control", Physica-Verlag, Heidelberg, 2001.

[17] F. Steimann, "On the use and usefulness of fuzzy sets in medical AI", Artificial Intelligence in Medicine, 21, 2001, pp. 131-137.

[18] N. Allahverdi, S. Torun, I. Saritas, "Design of a fuzzy expert system for determination of coronary heart disease risk", International Conference of Computer Systems and Technologies-CompSysTech'07.

[19] L. X. Wang, "Adaptive fuzzy systems and control", Prentice-Hall, Englewood Cliffs, NJ, 1994.

[20] R. J. Shalkoff, "Artificial neural networks ," Mcgraw-Hill, USA, 1997.

[21] Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and selection", 14thInternational Joint Conference on Artificial Intelligence,1995, 1137-1143 .

[22] G. Bontempi, M. Birattari, "Matlab software tool for neuro-fuzzy identification and data analysis," Technical report 1999. ftp://iridia.ulb.ac.be/pub/gbonte/public_software/fuzzy_manual.ps.Z.

[23] J. S. R. Jang, C. T. Sun, E. Mizutani, "Neuro-fuzzy and soft computing", Prentice-Hall International, New Jersey, 1997.

[24] www.mathwork.com/access/helpdesk/help/toolbox/fuzzy.

[25] J. S. R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference System", IEEE Trans. Systems, Man Cybern, 3, 2003, pp. 665–685.

[26] www.americanheart.org/Risk Factors and Coronary heart disease

[27] T. Gordon, W. B. Kannel, "Multiple risk function for predicting coronary heart disease", Am Heart. J., 103, 1982, pp. 1031-1039.

[28] T. Gordon, W. P. Castelli, M. C. Hjortland, W. B. Kannel, T. R. Dawber, "Diabetes, blood lipids, and the role of obesity in coronary heart disease risk for women", Ann Inten Med., 87, 1977, pp. 393-397.

[29] www.intelihealth.com/Risks and Prevention

[30] K. M. Anderson, P. W. Wilson, P. M. Odell, W.B. Kannel, "An updated coronary risk profile. A statement for health professionals", Circulation. J., 83, 1991, pp. 356–362.

[31] J. Delanghe, J. Robberch, M. De Buyzere, I. De Scheerder, O. Vanhaute, M. Beart et al, "Enzymatic creatine determination as early marker for myocardial infarction diagnosis", Fresenius' Journal of Analytical Chemistry, 330, 1988, pp. 366-367.

[32] J. Delanghe, M. De Buyzere, I. De Scheerde, L. Cluyse, H. Thierens, "Characteristic of creatine release during acute myocardial infarction, unstable angina and cardiac surgery", Clin. Chem., 41, 1995, pp. 928-933.

[33] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", second EDN, Kluwer, Boston, 1985.

[34] J. C. Bezdek, I. M. Anderson, "An application of the c-varieties clustering algorithms to polygonal cure fitting", IEEE Trans. Systems, Man Cybern, 15, 1985, pp. 637–641.