Original article:

Personalized healthcare monitoring using artificial neural network to Estimate 10 years risk of coronary heart disease

Haval Y. Yacoob Aldosky[1*], Amera Ismail Melhum[2]

[1]Physics Department, Faculty of Science, University of Duhok. Kurdistan Region, Iraq

[2]Computer Department, Faculty of Science, University of Duhok. Kurdistan Region, Iraq,

*Corresponding author: Haval Y. Yacoob Aldosky. Assist Prof ofMedical physics, PhD, Physics Department, Faculty of Science, University of Duhok. Kurdistan Region, Iraq

Abstract

Objectives: In recent years, there is an increasing demand for developing the personalized and non-hospital based care systems to improve the management of cardiac care. This study was aimed to develop an ANNtechnique for determining 10 years risk of coronary heart diseases (CHD) as, rapid, low cost method, that general population can easily use (home monitoring).

Materials and methods: A new ANN model was developed to estimate 10 years risk of coronary heart diseases for 208 women that mean age was (55±10.56 years). The subjects were classified to four group according to risk factors mainly systolic blood pressure and blood sugar. The predicted values of CHD risk estimated by the model were compared with those determined by Framingham equation and score of risks.

Results: Data analysis indicated that estimated 10 years CHD risks by ANN model showed statistically significant association with Framingham equation and score ($P < 0.001$). The mean correlation was 0.92 among the ANN model and Framingham equation and score risks for all groups.Based on the predicted and calculated risks, the majority of participants were classified at moderate risk (40 %) for 10 year CHD events.

Conclusion: ANN technology is highly promising in the development of accurate home diagnostic tools to evaluate the risk of coronary heart diseases for individuals and promote towards personalized healthcaremonitoring.To generalize the model, further studies should be done taking in account other region, ethnic and gender data samples.

Keywords: Artificial Neural Network, Coronary Cardiac Disease, Framingham Algorithm, personalized healthcare monitoring

Introduction

Artificial Intelligence (AI) is a study to simulate human intelligence into computertechnology. In point of view of benefits of AI in healthcare, many researchers are increasingly looking into new and innovative techniques with the help of information technology. Some of these advantages for AI, provide the necessary support to develop highly efficient automated diagnosticsystems and assisting doctors in making decision without consulting the specialists directly. This will provide the best health care facilities with low cost.

There are several AI tools implemented in medicine and health-related areas, for instance,Neural Networks, Fuzzy Logic, Support Vector Machines, Genetic Algorithms and Hybrid Systems. Neural Networks is nowadays the most promising area of interest. It is believed that for all the biomedical problems Neural Networks will prove to be the great solution in the coming years.

The application of artificial intelligence (AI) techniques in medicine has been grew since the development of high-speed digital computers. In reviewing this new field in 1984, Clancey and Shortliffe provided the following definition:'Medical artificial intelligence (MAI) is primarily concerned with the construction of AI programs that perform diagnosis and make therapy recommendations. Unlike medical applications based on other programming methods, such as purely statistical and probabilistic methods, medical AI programs are based on symbolic models of disease entities and their relationship to patient factors and clinical manifestations.[1]

These days, due to increase the mortality of patients who suffer from cardiovascular diseases (CVD) and the prevalence of sudden death, it became necessary to create an effective home alternatives accurate and quick to identify these risks by using personal electronic devices such as PC. This study was aimed to develop an AI software for determining 10 years risk of coronary heart diseases (CHD) as,rapid,low cost method, that general population can easily use (home monitoring) and to compare its performance against the Framingham Risk Equation (FRE) and Framingham Risk Score (FRS).

Theoretical aspects

This section introduce the basic concepts of CHD Risk Factors and Artificial Neural Networks ANN.

CHD Risk Factors

The main risk factors for CHD as reported in the literatures are age, sex, total cholesterol level, HDL, smoking status, hypertension, and preeclampsia and diabetes mellitus. The CHD risk factors included in this study were: age, gender, total cholesterol, high-density lipoprotein (HDL) cholesterol, arterial hypertension (Systolic), blood sugar, and current tobacco smoking status

Framingham Risk Equation

The Framingham Risk Equation is a predictive equation borne out of the Framingham Heart Study, which started in 1948 and has been operational for more than 60 years. The Framingham Risk Equation was developed for several cardiovascular disease endpoints by Anderson and colleagues in 1991.[2]The following variables were included in the equation: age, sex, systolic blood pressure, hypertensive, total and HDL-cholesterol values, and current smoking status (yes or no). The 10-year risk of CVD was classified as low (<10%), moderate (10% to 20%), or high (>20%).

Framingham Risk Scores

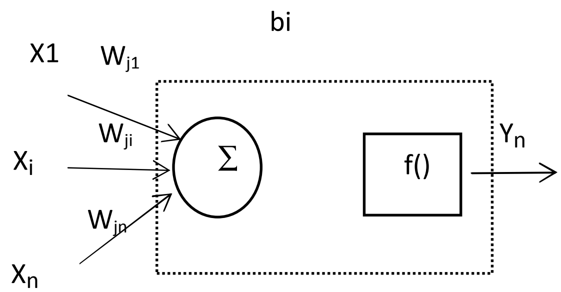
In 1998, the Framingham risk score for estimating 10-year total CHD events for asymptomatic individuals was used to estimate CHD risk. FRS was derived from a largely Caucasian population of European descent. FRS included age, total cholesterol, HDL cholesterol, blood pressure, diabetes, and smoking status.[3]Subsequent studies have suggested that the Framinghamrisk score performs well for prediction of CHD events in black and white women and men.[4]Points are added together to predict the 10-year risk score of Myocardial Infarction. Specific scores are given to each factors. These scores are summed to obtain the Risk factor

Artificial Neural Network

An Artificial neural Network (ANN) is a computational model which imitates the human brain process. The ANN consists of a set of processing nodes which called neuron, these neuron are connected to each other, that operating in parallel, these neuron are connected together to construct the layers, as in figure 1.1 each node receive its input values from input vector Xi and sums up its after multiplying with respective connections Wij then add the bias bi to them, after that apply the activation function for the result which be the value of output Yn,

$$Yi = f\left(\sum Xiwij + bi \right)$$

The network consist of an input and output layers between these two layers there are one or more than one layers called hidden layers as shown in Fig 1.2. For constructing ANN there is no rule to select the number of nodes in each hidden layer and the number of hidden layers, it's done by trial and error.[5, 6]



Proposed Method

Basically, there are five major steps involved in the methodology as below:

1- Data Collection

This work performed a descriptive, cross-sectional study which involved 208women that the mean age was (55±10.56 years). Subject characteristics and risk factors were collected from General Central Laboratory in Duhok City). Subject characteristics involved body weight and height which were measured to the nearest 0.1 Kg using a digital weighing scale, and 0.1 cm, using a graduated elastic tape, respectively. Body mass indexwasderived from the weight and height of an individual. While the risk factors included blood serum total cholesterols in mg/dl, high density lipid (HDL) cholesterols, fasting blood sugar in mg/dl, and systolic blood pressure in mmHg. For the smoking status, subjects who are currently smoking were coded to 1 (smoking) while subjects who are currently non-smoking, were coded to 0 (non-smoking).Table 1 shows the baseline clinical characteristics of the subjects.

Table 1. The characteristics (Risk Factors) of the subjects

Characteristics	Median± SD	Minimum	Maximum
Age (years)	**55±10.56**	**32**	**85**
Systolic Blood Pressure(mm Hg)	**140±24.54**	**80**	**200**
Total cholesterol (mg/dL)	**192±49.1**	**79**	**412**
HDL cholesterol (mg/dL)	**33.2±8.68**	**15.7**	**73**
Blood sugar (mg/dL)	**117±67.81**	**42**	**409**
Height (cm)	**151±6.4**	**134**	**165**
Weight(Kg)	**73.3±14.28**	**43**	**132**
BMI	**31.7±5.58**	**19.8**	**53.2**
Smoking	**No**		

2- Data Preprocessing

The 10 year risk of coronary heart diseases was calculated for four different categorizes groups according to two risk factors systolic blood pressure and blood sugar. The first group (61 subjects) based on normal systolic blood pressure and blood sugar. The second group (65 subjects) based on normal systolic blood pressure but high blood sugar. In opposite, the third group (40 subjects) based on normal blood sugar level with high systolic blood pressure. Finally, the last group (42 subjects)based on high both systolic blood pressure and blood sugar level. The normal range for systolic blood pressure and blood sugar were determined by <140 mmHg and <110 mg/mL respectively. In each case, the remaining risk factors entered in to analysis were age, sex, total cholesterols, high density lipid, and smoking status coded (0,1).Subjects with group 1 and 2 were more than that of subjects with group 3 and 4.

Table (2) Prevalence of risk within subject categories for CHDstratified by risk percentage

Risk	No. of subject	Mean risk in category (%)	Percent of all subjects
No Risk<1%	**1**	**0.87**	**0.48%**
Low Risk<10%	**73**	**6.51**	**35.09%**
Medium Risk10-20%	**81**	**14.46**	**38.94%**
High Risk>20%	**53**	**24.67**	**25.48%**

www.ijbamr.com P ISSN: 2250-284X , E ISSN : 2250-2858

Table (3) Prevalence of risk within subject categories for CHDstratified by age.

Age/years	**Risk Level**			
	No risk	**Low**	**Medium**	**High**
30-40	**0**	**5**	**8**	**4**
41-50	**0**	**28**	**23**	**15**
51-60	**1**	**16**	**32**	**18**
61-70	**0**	**21**	**12**	**8**
>71	**0**	**3**	**6**	**8**
Total	*1*	*73*	*81*	*53*

3- Artificial Neural Network Model

In this paper back propagation algorithm has been used as the training algorithm which is a type of the multilayer feed-forward neural network to predict the risk value of coronary heart diseases (CHD), our goal is train ANN to estimate the 10 year risk of CHD based on age, smoking, blood sugar, ECG, HDL , SBP and total cholesterol. Data was provided by General Central Laboratory in Duhok City. The numbers of dataset are 208 samples for women, 70% has been used for training and 30%for testing.

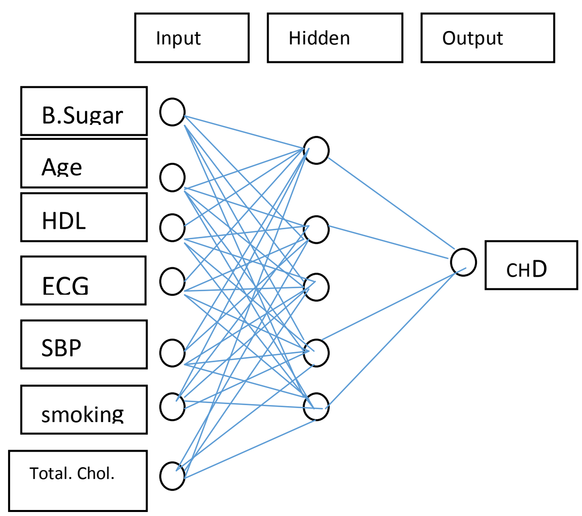


Figure (2) Neural Network Structure

The main step for training algorithm for BPN is as follows

Step0. Initialize weights and biases. (Set to small random values)

Step1. Training the each pair of data, do Steps 2-5.

Step2. Vector X represent input unit $(X_i,\ i = 1 \dots n)$ receive input data X_i and send it to next layer (the hidden units).

Step3. Compute the output for each hidden unit $\left(Z_j,\ j = 1 \dots p\right)$

$$Z_j = f(v_{oi} + \sum_{i=1}^{n} x_i\, v_{ij})$$

Step4. Compute the value for each output unit $(Y_k,\ k = 1 \dots m)$

$$y_k = f(w_{0k} + \sum_{j=1}^{p} z_j\, w_{jk})$$

Step5. Each output unit $(Y_k,\ k = 1 \dots m)$ adjust its bias and weights

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$$

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$$

In which Wjk corresponding to the vector of weights and Vij to the biases. At the end of training phase of the neural network, the weights of neural network are saved for using later in testing phase

4- **Statistical methods**

Summary values are given as mean ± SD or median for risk factors values.The ability of ANN model to estimate coronary heart disease risk was evaluated by simple linear regression analysis. A Pearson correlation coefficients were calculated which tells us how strong the linear relationship is. R-square was also calculated which is a coefficient of Determination which tells us how many points fall on the regression line. R-square value was determined graphically by plotting the observed Vis estimated 10-year risk as well. The reliability (statistically significant) of results was tested by Significance F and P-values. For all analyses, p values less than 0.05 were considered statistically significant. Analyseswere performed using Excel 2013.

Results

Table (1) shows208 women that the mean age was (55±10.56 years). Some of them (nearly by 70) had none or only one risk factor. The prevalence of both hypertension and diabetes was similar to that of the general population. Mean level of total cholesterol was within normal range, whereas the mean HDL cholesterol level was within (33.2±8.68) which is considered low range. Tables (2 and 3) show the prevalence of risk within subject categories for CHD stratified by risk percentage and age respectively.

Table 4 presents the statistical parameters R^2 (percentage of explained variation) and ρ (correlation coefficient) for FRE and FRS one a time with ANN for all the groups. The best resultsof R^2 and ρ (0.999) were achieved in second group (high SBP and normal BS) with ANN model. In contrast, lower values of these statistical parameters were shown (0.64, 0.812) in fourth group (high SBP and high BS) correlated with equation, while by comparing with score these parameters were (0.671, 0.833). There was no significantly difference (P< 0.001)of CHD risks predicted by ANN model comparing with both FRE and FRS.

Graphically, figures 3 (a,b,c and d) depict the predicted values of CHD risksestimated by ANN model were showed good agreement (around R^2= 0.90) with calculated values of CHD risk from both FRE and FRS for first three first group, while less agreement (R^2= 0.67) was shown for the fourth group. The optimal agreement (R^2= 0.99) was achieved for ANN model in second group. In all groups, the correlation was statistically acceptable with high significant (P<0.001).

Figure (4) presents the estimated CHD risk of the all subjects. Based on the ANN model, the majority of subjects were classified at moderate risk (40 %) for 10 year CHD events. The percentage of subjects classified as having a high risk of 10 year CHD about 28%. Based on Framingham Risk Equation, 35 % of the participants were classified as low-risk, 39 % as moderate risk, and 25 % as high-risk patients for

CHD.Framingham Risk Score, 34 % of the participants were classified as low-risk, 38 % as moderate risk, and 28 % as high-risk patients for CHD. This result is also compatible with Framingham equation and score as well.

Table (4) Comparing Framingham equation and score with ANN for estimating 10 year risk of CHD.

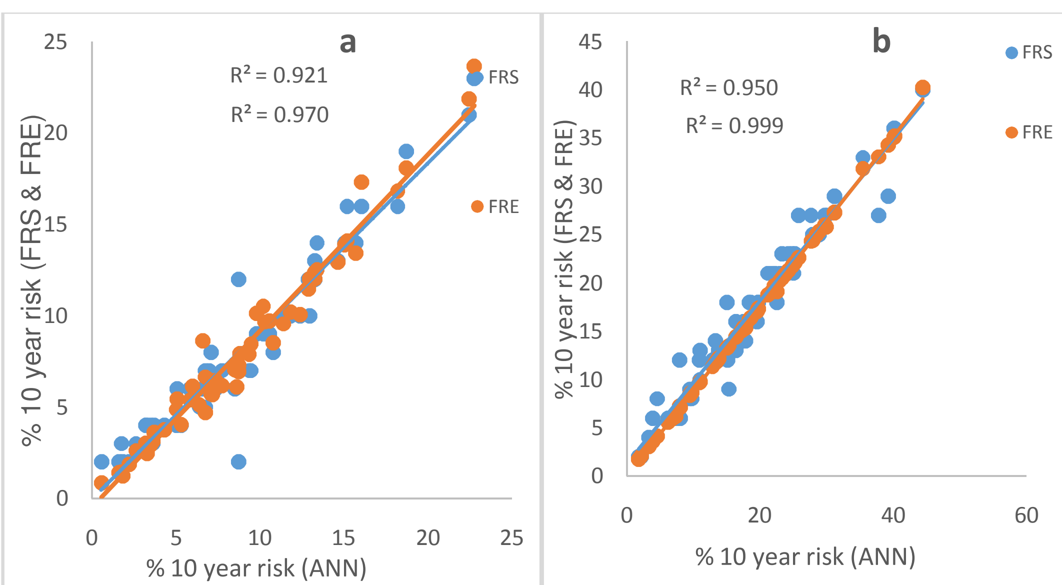
Method	ANN											
	G1			G2			G3			G4		
	R^2	P	P-value	R^2	ρ	P-value	R^2	ρ	P-value	R^2	P	P-value
Framingham equation	0.954	0.974	<0.001**	0.999	0.999	<0.001**	0.806	0.897	<0.001**	0.64	0.812	<0.001**
Framingham Score	0.903	0.950	<0.001**	0.950	0.974	<0.001**	0.848	0.921	<0.001**	0.671	0.833	0.001**

G1- Normal systolic blood pressure& Normal blood sugar, G2- Normal systolic blood pressure & abnormal blood sugar, G3-Abnormal systolic blood pressure & Normal blood sugar, G4-Abnormal systolic blood pressure & abnormal blood sugar

ρ= correlation coefficient

* P< 0.05significant

** P< 0.001 highly significant.



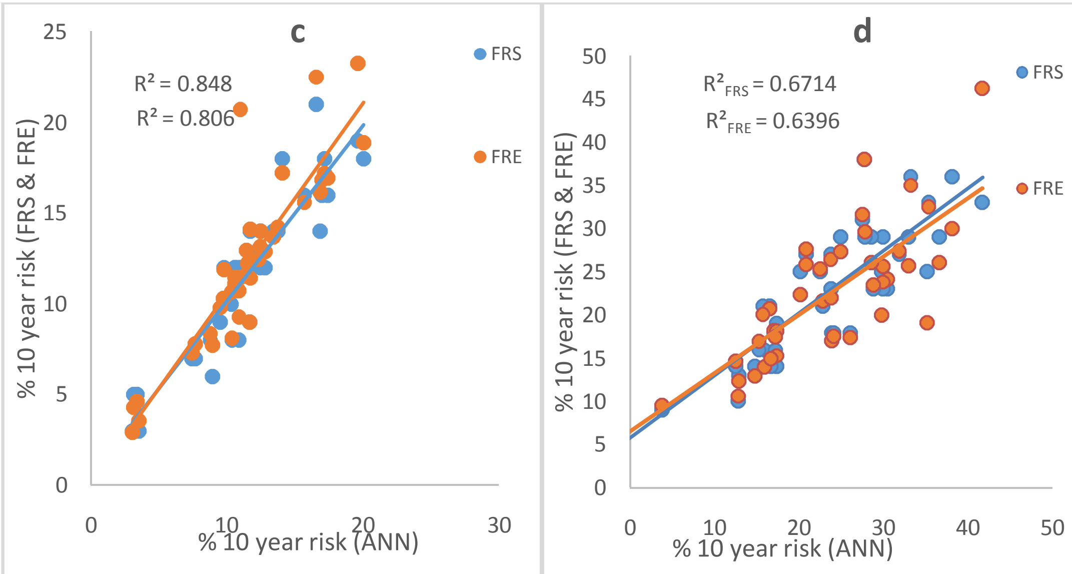


Figure 3. Correlation between predicted values of 10 years risk of CHD by ANN model with calculated values from Framingham risk equation and score for (a) first group (normal blood pressure andnormal blood sugar), (b) second group (normal blood pressure and high blood sugar), (c) third group (high blood pressure and normal blood sugar) and (d) fourth group (high blood pressure and high blood sugar).

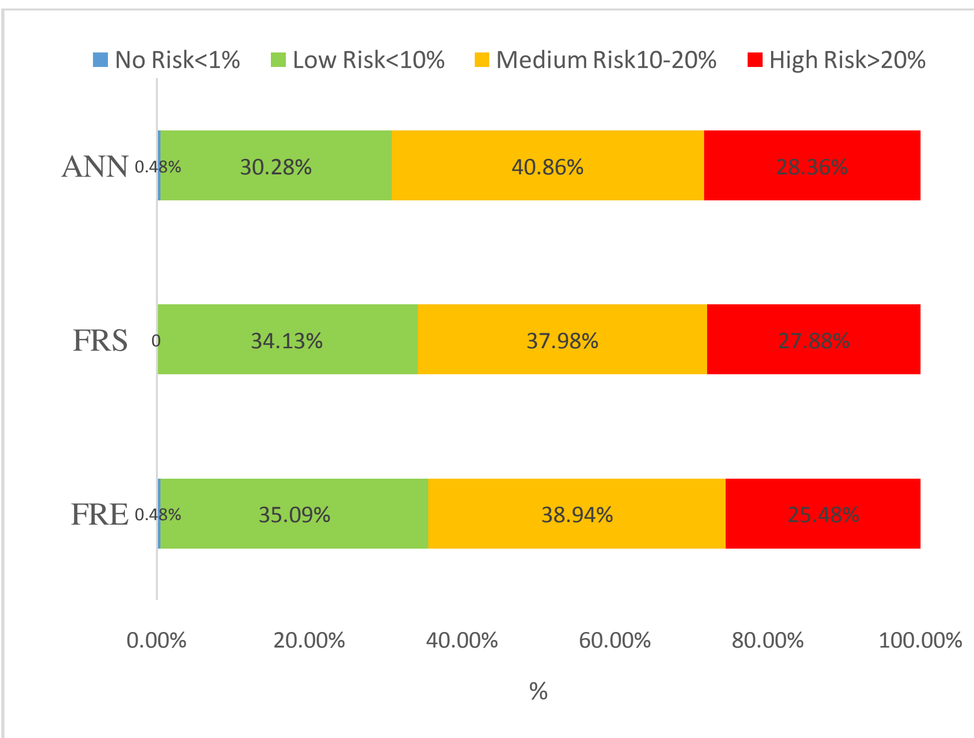


Figure 4: Prevalence of estimated CHD risks according to ANN and calculated by Framingham Risk Score, Framingham Risk Equation with aggravating percentage for subjects.

Discussion

Nowadays, the sophisticated technology managed the old idea which stated that the techniques including cluster analysis and neural networks are difficult to apply in clinical practice through a simple, low cost, high accuracy and more importantly, the various combinations of the risk factors. Ultimately, a simple format which preferred by clinicians [7, 8]such as the Framingham scoring sheets [9]and the colored SCORE charts. The advantage of cardiovascular risk calculator is to estimate the risk of a cardiovascular event occurring over a given time period. The favorite method for most people is Framingham equation. This because of its advantage of allowing calculations over various time periods (4 to 12 years) and for different outcomes: cardiovascular disease, stroke, coronary disease and myocardial infarction.

The results of this study showed that the ability of ANN technique in determining 10 years of CHD risks was reasonable and comparable with other methods. It is evident from Figure (3) that the correlation coefficient decreases whenever the risk factors mainly SBP and BS directed toward abnormalities in their range. This may be solved by considering other parameters such as whether or not the person taking medicine for these abnormalities.Consequently, the accuracy of ANN model can be improved not only by increasing the number of risk factors, but also by their precise selection. However, reasonable accuracy in predicting CHD and its prevalence among the subjects has been demonstrated when compared with calculating CHD by Framingham equation and Framingham score applied to the subjects.In addition, this simple, low-cost, and rapid ANN model will be useful for the detection of 10 years CHD risks at home.

The level of prediction accuracies of ANN model could be much higher when a much larger dataset from a wider variety of hospitals across the globe can be considered.

Limitations

The study has three limitations. First, the sample size of 208 (145 for training, 63 for testing) might be small for a model test. Second, this is a cross sectional study. These two limitations are due to lack of health insurance system. Third, further studies should be done taking in accountother region, ethnic and gender data samples.

Future work

According to industry estimates, nowadays, 500 million smartphone users worldwide have used a health care application and by 2018, 50 percent of the more than 3.4 billion smartphone and tablet users will have downloaded mobile health applications.[10]The sensible future goal of this project willbe directed toward creating a software algorithms combined with smartphone technology to provide widespread healthcare services and as a potential to reduce costs and errors.This software allows users to know 10 years risk of coronary heart disease by entering data such as their age, sex, total cholesterol level, HDL, smoking status,systolic and diastolic blood pressure,and blood glucose concentration into their smartphones.

References

[1] William J. Clancey and Edward H. Shortliffe [Readings in Medical Artificial Intelligence: The First Decade] Addison Wesley, Reading, MA, 1984

[2] Anderson K M, P W Wilson, P M Odell and W B Kannel [An updated coronary risk profile. A statement for health professionals] Circulation. 1991;83:356-362

[3] Peter W.F. Wilson, MD; Ralph B. D'Agostino, PhD; Daniel Levy, MD; Albert M. Belanger, BS; Halit Silbershatz, PhD; William B. Kannel, MD [Prediction of Coronary Heart Disease Using Risk Factor Categories] Circulation. 1998;97:1837-1847.

[4] D'Agostino RB Sr, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001; 286:180.

[5] Faustett L., "Fundamentals of Neural Networks: Architecture Algorithms and Application", 1st edition, 1993.

[6] Ching. C, "Application of Back-propagation Neural Network to Categorization of Physical Fitness Levels of Taiwanese Females", Journal of Medical and BiologicalEngineering,2011,31(1).

[7] Isles CG, Ritchie LD, Murchie P, Norrie J. Risk assessment in primary prevention of coronary heart disease: randomised comparison of three scoring methods. BMJ. 2000;320:690–691.

[8] Graham IM, Stewart M, Hertog MG. Factors impeding the implementation of cardiovascular prevention guidelines: findings from a survey conducted by the European Society of Cardiology. Eur J Cardiovasc Prev Rehabil. 2006;13:839–845.

[9] Copie X, Hnatkova K, Staunton A. Predictive power of increased heart rate versus depressed left ventricular ejection fraction and heart rate variability for risk stratification after myocardial infarction: results of a two-year follow-up study. J Am Coll Cardiol. 1996;27:270 –276.

[10] Jahns Ralf-Gordon. 500m people will be using healthcare mobile applications in 2015. 10 November 2010. Available from (http://www.research2guidance.com/500m-people-will -be-using-healthcare-mobile-applications-in-2015/).