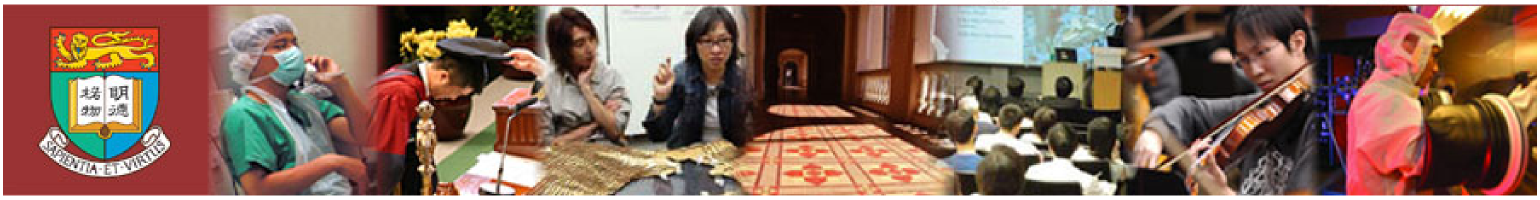
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by HKU Scholars Hub

The HKU Scholars Hub The University of Hong Kong 香港大學學術庫



Title	**Bio-medical application on predicting systolic blood pressure using neural networks**
Author(s)	**Wu, TH; Kwong, EWY; Pang, GKH**
Citation	**The 2015 IEEE 1st International Conference on Big Data Computing Service and Applications (BigDataService), Redwood City, CA., 30 March-2 April 2015. In Conference Proceedings, 2015, p. 456-461**
Issued Date	**2015**
URL	**http://hdl.handle.net/10722/214833**
Rights	**IEEE International Conference on Big Data Computing Service and Applications (BigDataService). Copyright © IEEE Computer Society.**

2015 IEEE First International Conference on Big Data Computing Service and Applications

Bio-medical Application on Predicting Systolic Blood Pressure Using Neural Networks

Tony Hao WU
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong
Email: haowu@eee.hku.hk

Enid Wai-Yung KWONG
School of Nursing
The Hong Kong Polytechnic University
Hong Kong.
Email: enid.kwong@polyu.edu.hk

Grantham Kwok-Hung PANG
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong
Email: gpang@eee.hku.hk

***Abstract*— This paper presents a new study based on artificial neural network, which is a typical technique for processing big data, for the prediction of systolic blood pressure by correlated factors (gender, serum cholesterol, fasting blood sugar and electrocardiographic signal). Two neural network algorithms, back-propagation neural network and radial basis function network, are used to construct and validate the bio-medical prediction system. The database of raw data is divided into two parts: 80% for training the neural network and the remaining 20% for testing the performance. The experimental result shows that artificial neural networks are suitable for modeling and predicting systolic blood pressure. This novel method of predicting systolic blood pressure contributes to giving early warnings to adults who may not take regular blood pressure measurements. Also, as it is known that an isolated blood pressure measurement is sometimes not very accurate due to the daily fluctuation, our predictor can provide another reference value to the medical staff.**

Keywords*— *Systolic blood pressure; hypertension; bio-medical; big data application; machine learning.

I. Introduction

It is known that big data technology could be applied in providing valuable insights about people's condition, such as monitoring health behaviors, indicting potential factor to disease, and giving suggestion to people and medical staffs [1][2].

Hypertension is a major global public health issue. According to the National Institutes of Health (NIH), around one in three adults in the US have hypertension [3]. It would be a chronic medical condition if a person is characterized by persistent high blood pressure (BP) with systolic and diastolic BP readings of higher than 140mmHg and 90 mmHg respectively [4].

It is desirable to develop a device which can provide a prediction of the blood pressure based on health factors (serum cholesterol, fasting blood sugar and electrocardiographic signal) of a person. The method would enable us to know the risk of high blood pressure. This paper aims to develop a method to predict the systolic blood pressure of a person given the factors that have an association with blood pressure. The developed method does not aim to replace the need to measure blood pressure of a person directly, and its clinical relevance will be explained later.

It must be noted that the normal blood pressure range of a person should correspond to his age group [5]. It is well-known that blood pressure has daily pattern with various BP readings within a day [6]. The blood pressures are usually lower when asleep, and it rises during the day and reaches a peak value in mid-afternoon. Then, blood pressure begins to drop again and during sleep, blood pressure could be 20% lower than during the day.

The clinical relevance of the developed method is explained as follows. Suppose the blood pressure of a person is not available on hand or the person is at a remote location. Figure 1 shows an artificial neural network that can be used to predict the systolic blood pressure of a person based on a database used in this study. In this example, the person has to input some values related to age, gender, serum cholesterol, fasting blood sugar and electrocardiographic signal. The neural network would give a prediction on his systolic blood pressure. Assuming that the predicted value is larger than the hypertension alarm value, warning information should be provided to the user. For example, the predicted value is 150mmHg, which is more than the systolic BP of 142mmHg for a person of aged at 52, alerts should be given.

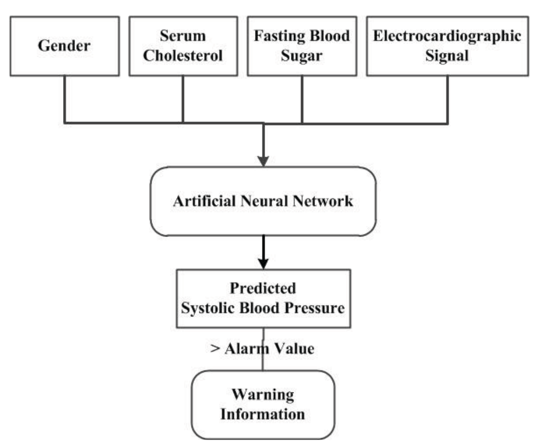


Fig. 1 - Predicting without Measurement

978-1-4799-8128-1/15 $31.00 © 2015 IEEE
DOI 10.1109/BigDataService.2015.54

IEEE computer society

Another scenario on the usefulness of the blood pressure predictor is as follows. A person provides all the relevant information for the neural-network-based blood pressure predictor and his blood pressure measurement is also taken (Fig. 2). Suppose the measured value seems normal, but the predicted value is at an alarming value, the person would be recommended to recheck on his blood pressure. For example, the measured value is 138mmHg for a person of aged 52, but the predictor gives a value of 150mmHg. The relatively low value of his systolic blood pressure measurement could be inaccurate due to the daily frustration or an unknown reason. In such a situation, the person should double check and is advised to have further measurements of his blood pressure.

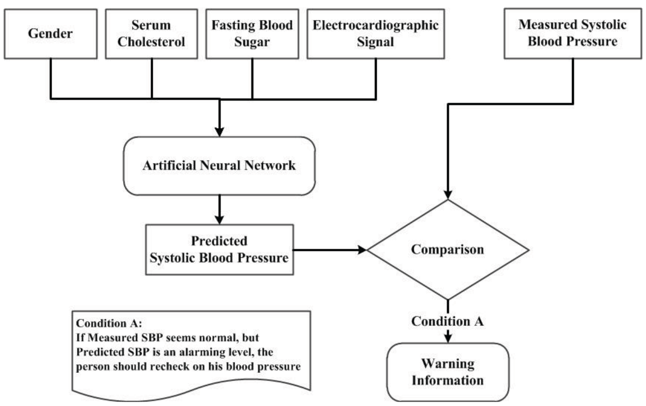


Fig. 2 - Predicting with Measurement

In this paper, the bio-medical application is realized by the machine learning method of artificial neural network, which will learn the relationship between the correlated variables (serum cholesterol, blood sugar and electrocardiographic signal etc.) and the systolic blood pressure value. A trained neural network would then be used to predict the systolic blood pressure when the input variables are provided.

II. CORRELATION OF BLOOD PRESSURE AND OTHER VARIABLES

According to World Health Organization [3], high blood pressure is very often associated with some health factors of a person such as old age, gender, high cholesterol level, danger blood glucose and electrocardiography. Below is a discussion of medical factors associated with high blood pressure.

A. Age

Sparrow et al. [7] used the multiple linear regression analyses to indicate that the systolic blood pressure (SBP) of a person aged over 50 years old would increase more rapidly than the SBP of those whose age was between 20 and 39.

B. Male/ Female

According to a survey [7], blood pressure of male is usually higher than that of the female at the same age. In fact, men are in higher risk for hypertension and cardiovascular disease than premenopausal women of the same age. It also shows that blood pressure goes up with age for both male and female. In addition, men have higher daily mean blood pressure than women until the age has reached 80.

C. Serum cholesterol

According to Felorey et al. [8], blood pressure and serum cholesterol are indeed dependent on each other and should therefore be expected to have effects on a series of heart disease. Another study by Hjermann et al. [9] revealed that body mass index and serum triglycerides would influence the relationship between blood pressure and cholesterol significantly.

D. Blood sugar/ glucose

Blood sugar concentration or blood glucose level indicates the source of energy for the body's cells and blood lipid. Conway et al. [10] studied the relationship between the magnesium ion present in blood plasma and systolic blood pressure using ANN. In addition, ANN is also used to find the relationship between the magnesium ion present in blood plasma and blood glucose.

E. Electrocardiography

The electrocardiography (ECG) records the activity condition of the heart, which detects and amplifies the electrical impulse signal from different directions during each heartbeat.

III. MATERIAL AND METHODS

A. Artificial neural network

Artificial neural network (ANN) could be applied in big data analytics, which is based on intelligent computational model, and uses the computer network system to simulate the biological neural network. A neural network consists of an input layer of source, at least one middle or hidden layer of computational neurons, and an output layer of computational neurons [11, 12, 13]. The ability of an ANN is to discover the nonlinear relationship between its inputs and outputs. The neural network technology has been used as a major role in many mathematic, scientific and industrial applications, such as time series prediction, pattern recognition, decision making, event prediction etc. [14, 15]. In this paper, the neural network is trained to learn the correlation between the input variables and the output, which is the systolic blood pressure. Essentially, we would like to use an artificial neural network as a tool to predict the value of the blood pressure of a person, given some health-related measurements and data of a person.

For example, Fig. 3 shows the ANN structures using the database in this study. The ANN structure gives a prediction structure based on four inputs (age, cholesterol, blood sugar and ECG) and three hidden nodes. The objective of the structure is to predict the systolic blood pressure based on the different input variables.

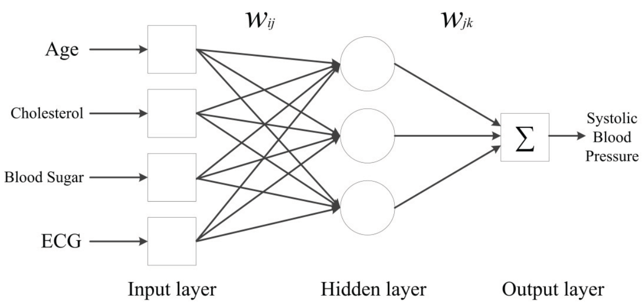


Fig. 3 - Structure of Artificial Neural Network

The strength of a neural network is on its learning capability. Two well-known algorithms for the training of neural networks are back-propagation (BP) algorithm, and the radial basis function (RBF) network. The interconnection weights between the different layers in an ANN will be obtained at the end of the training.

B. Back-propagation Neural Network [11, 12, 16, 17]

The back-propagation (BP) neural network is a typical architecture of multi-layered feed-forward neural network. A typical BP neural network is usually consisted of layers of neurons, and the objective is to train the network weights so as to minimize the mean-square error of the network output. The architecture of a typical BP neural network is composed of four inputs, one hidden layer with three hidden nodes, and one output value.

C. Radial Basis Function Neural Network [18, 19]

Our radial basis function (RBF) neural network consists of one hidden layer. In a RBF network, the hidden layer can implement a better nonlinear mapping from the data space to the feature space than BP neural network. The most common function chosen is a Gaussian function, which means that the closer the input to the center of the Gaussian, the larger the response of the node. In stage one of the training of the RBF network, the self-organizing feature map (SOFM) is used as a clustering algorithm to determine the weights between the input and the hidden layer. In the second stage, the weights between the hidden and output layer are determined by a least-mean-square (LMS) algorithm. The convergence of the weights is found to be quick and efficient.

D. Datasets

The heart disease dataset were obtained from the widely used bio-medical database stored at the UCI machine learning repository [20]. This database contained 14 attributes, but only a subset of six attributes was used in the experiment. In particular, the attributes used are age (vary from 29 to 65), resting blood pressure, serum cholesterol (in mg/dl), fasting blood sugar, resting electrocardiographic results and maximum heart rate achieved. The database can be divided into two subsets: male dataset (165 cases) and female dataset (75 cases).

IV. RESULTS

The experiments were carried out by two different algorithms of artificial neural networks: the back-propagation (BP) and radial basis function (RBF). The BP algorithm is first used for the training of an ANN. The output is the predicted value of the systolic blood pressure. The number of inputs is the factors used for the prediction, and the number of hidden nodes varies from around 2 to 10.

The RBF algorithm is also used for the training an ANN, which would map the relationships between systolic blood pressure (SBP) and input variables belonging to different clusters. The number of clusters naturally formed is determined first, which would provide the weights between the input and hidden layer. In the second stage, the weights are determined by an iterative least-mean-square (LMS) algorithm.

It is very important to emphasize that the testing data is separate from the training data. Given the data from a database, a percentage (80%, 85% or 90%) is used for training, and the rest is for testing the performance of the trained neural network. The absolute difference (error) between the real value and predicted values will be recorded. Each result is based on the averages of twenty-five cases of training and testing.

A. Back-propagation ANN

Male

For the database that was used for developing a systolic blood pressure predictor, different percentages of the database have been used for the training of the BP ANN. It is found that the best result is obtained when 80% is used for training, and the remaining 20% used for testing the performance of the ANN (Table 1).

Table 1 - Experimental results based on male case using BP ANN

	error less than 5	error less than 10	error less than 15	error less than 20	error less than 25	error less than 30
Percentage of data for training (3 hidden nodes, inputs are Age, Cholesterol, Blood Sugar, ECG)						
80%	**27.9%**	**49.9%**	**68.6%**	**81.0%**	**90.6%**	**96.1%**
85%	22.1%	42.6%	65.5%	82.8%	93.2%	96.5%
90%	26.2%	47.8%	71.7%	87.5%	95.5%	97.2%
Number of hidden nodes (80% of data for training, inputs are Age, Cholesterol, Blood Sugar, ECG)						
2	27.2%	47.9%	69.6%	86.0%	94.2%	97.0%
3	**27.9%**	**49.9%**	**68.6%**	**81.0%**	**90.6%**	**96.1%**
4	24.3%	49.5%	68.3%	84.1%	92.7%	95.6%
5	24.0%	47.9%	67.0%	81.4%	89.7%	94.0%
6	21.8%	44.9%	62.6%	76.0%	84.0%	91.1%
10	20.3%	41.0%	59.3%	70.5%	79.4%	85.5%

Different inputs (3 hidden nodes, 80% of data for training)						
Age, Cholesterol, Blood Sugar, ECG	**27.9%**	**49.9%**	**68.6%**	**81.0%**	**90.6%**	**96.1%**
Age, ECG, Blood Sugar, Heart Rate	28.8%	49.7%	70.0%	86.8%	94.8%	98.6%
Age, Cholesterol, ECG, Heart Rate	22.5%	49.3%	65.3%	77.9%	89.7%	96.4%
Age, Cholesterol, Blood Sugar, Heart Rate	25.3%	47.5%	66.2%	79.2%	89.5%	94.5%
Age, Cholesterol, Blood Sugar, ECG, Heart Rate	24.2%	47.0%	64.7%	82.6%	92.2%	95.7%

Error: absolute difference between predicted and actual SBP (in unit of mmHg)

Female

For the female case, as shown in Table 2, it seems that the use of 3 hidden nodes, with 80% of data for training is giving the "best" result. Similar to the case of male, the use of additional inputs does not help with the prediction.

Table 2 - Experimental results based on female case using BP ANN

	error less than 5	error less than 10	error less than 15	error less than 20	error less than 25	error less than 30
Percentage of data for training (3 hidden nodes, inputs are Age, Cholesterol, Blood Sugar, ECG)						
80%	**33.2%**	**57.6%**	**76.0%**	**82.4%**	**87.8%**	**92.2%**
85%	29.3%	55.0%	71.4%	83.4%	88.4%	98.2%
90%	14.6%	38.7%	62.1%	74.6%	82.0%	87.5%
Number of hidden nodes (80% of data for training, inputs are Age, Cholesterol, Blood Sugar, ECG)						
2	29.2%	51.6%	65.9%	78.4%	86.8%	92.8%
3	**33.2%**	**57.6%**	**76.0%**	**82.4%**	**87.8%**	**92.2%**
4	27.5%	51.8%	64.4%	74.1%	80.9%	87.6%
5	20.4%	40.8%	55.7%	66.2%	74.0%	80.3%
6	22.2%	39.7%	56.0%	69.1%	76.6%	80.6%
10	18.0%	31.1%	41.1%	53.8%	60.9%	66.0%
Different inputs (3 hidden nodes, 80% of data for training)						
Age, Cholesterol, Blood Sugar, ECG	**33.2%**	**57.6%**	**76.0%**	**82.4%**	**87.8%**	**92.2%**
Age, ECG, Blood Sugar, Heart Rate	24.9%	54.8%	68.1%	74.7%	83.4%	88.9%
Age, Cholesterol, ECG, Heart Rate	27.2%	51.7%	69.4%	81.3%	87.4%	93.3%
Age, Cholesterol, Blood Sugar, Heart Rate	25.2%	41.0%	57.5%	78.8%	77.4%	87.0%
Age, Cholesterol, Blood Sugar, ECG, Heart Rate	24.8%	45.7%	63.2%	73.6%	80.7%	87.1%

Error: absolute difference between predicted and actual SBP (in unit of mmHg)

B. Radial-Basis Function (RBF) Network

Male

In this experiment, 165 male cases are selected from the database. From Table 3, with the use of different percentages (80%, 85%, 90%) of data for training the system, 80% for training seems to be the best choice. Also, more variables/factors do not seem to improve on the prediction.

Table 3 - Experimental results based on male case using RBF ANN

	error less than 5	error less than 10	error less than 15	error less than 20	error less than 25	error less than 30
Percentage of data for training (4 RBF nodes, inputs are Age, Cholesterol, Blood Sugar, ECG)						
80%	**26.2%**	**51.3%**	**69.7%**	**81.3%**	**93.8%**	**97.8%**
85%	27.4%	49.2%	68.6%	80.5%	91.4%	96.8%
90%	28.9%	48.9%	69.7%	84.1%	92.5%	97.6%
Different number of inputs (4 RBF nodes, 80% of data for training)						
Age, Cholesterol, Blood Sugar, ECG	**26.2%**	**51.3%**	**69.7%**	**81.3%**	**93.8%**	**97.8%**
Age, ECG, Blood Sugar, Heart Rate	25.3%	51.3%	70.7%	84.4%	93.9%	98.1%
Age, Cholesterol, ECG, Heart Rate	25.8%	50.2%	68.4%	83.8%	94.2%	98.4%
Age, Cholesterol, Blood Sugar, Heart Rate	27.3%	51.1%	70.2%	82.8%	93.5%	98.0%
Age, Cholesterol, Blood Sugar, ECG, Heart Rate	26.9%	51.3%	68.9%	83.2%	93.7%	97.9%

Error: absolute difference between predicted and actual SBP (in unit of mmHg)

Female

In this experiment, 75 female cases are selected from the database. Again, the use of 80% for training seems to be the best choice (Table 4). The use of an additional input does not seem to provide any better result.

Table 4 - Experimental results based on female case using RBF ANN

	error less than 5	error less than 10	error less than 15	error less than 20	error less than 25	error less than 30
Percentage of data for training (4 RBF nodes, inputs are Age, Cholesterol, Blood Sugar, ECG)						
80%	**26.2%**	**51.3%**	**69.7%**	**81.3%**	**93.8%**	**97.8%**
85%	27.4%	49.2%	68.6%	80.5%	91.4%	96.8%
90%	28.9%	48.9%	69.7%	84.1%	92.5%	97.6%
Different number of inputs 4 RBF nodes, 80% of data for training)						
Age, Cholesterol, Blood Sugar, ECG	**26.2%**	**51.3%**	**69.7%**	**81.3%**	**93.8%**	**97.8%**
Age, ECG, Blood Sugar, Heart Rate	25.3%	51.3%	70.7%	84.4%	93.9%	98.1%
Age, Cholesterol, ECG, Heart Rate	25.8%	50.2%	68.4%	83.8%	94.2%	98.4%
Age, Cholesterol, Blood Sugar, Heart Rate	27.3%	51.1%	70.2%	82.8%	93.5%	98.0%
Age, Cholesterol, Blood Sugar, ECG, Heart Rate	26.9%	51.3%	68.9%	83.2%	93.7%	97.9%

Error: absolute difference between predicted and actual SBP (in unit of mmHg)

V. Discussion

We have developed separate ANNs for both male and female. For the given database, the ages of the subjects with 165 male cases and 75 female cases.

Many results have been presented in the previous section. The reliability of the proposed approach was evaluated by comparing the ANN-predicted systolic blood pressure value against the data not used for training but reserved for evaluating the performance of the trained ANN. The result of the comparisons between the two values has shown that in around 30% of the cases, the error in the prediction is within 5 mmHg. Also, in around 50% of the cases, the error in the prediction is within 10 mmHg. The prediction error can be considered to be acceptable. It is believed that if a larger database of training cases can be provided, the error of prediction can be further reduced to within a few mmHg with a high level of confidence. This can be a direction for future work.

In an evaluation on the use of different percentages of data for training, in all the cases observed, the use of 80% of the data in the database for training, with the remaining 20% for testing the performance of the neural network, seems to give the best performance. Although we have not tested for all possible percentages, this ratio of 80-20 is also a heuristic commonly used in the field of neural network. Our results verify the validity of this heuristic.

The structure of both artificial neural networks (ANN) studied in this paper has only one hidden layer. In fact, we have also attempted the use of two hidden layers in the back-propagation ANN but the results are all worse than one hidden layer.

It is also known in the field of ANN that one hidden layer is sufficient to represent any nonlinear function provided enough hidden nodes are used. We have also found that the use of only a few hidden nodes (three to five hidden nodes) seems to give the best performance. We have attempted the use of more hidden nodes (e.g. ten or even more hidden nodes) in the training of both BP and RBF networks, but the performance of the prediction is inferior when compared with the use of just a few hidden nodes. This observation is not surprising as it is well-known in the field of neural networks that the use of more hidden nodes would just introduce more parameters into the predictor (ANN). The neural network would try to "memorize" the training cases rather than to generalize and strengthen is learning capability. This issue of over-training should be avoided and the use of just a few hidden nodes has verified that our approach is correct.

Comparing between BP and RBF networks in the current study, the results seem to show that they are comparable. The BP structure is providing a default ANN configuration and the learning effort goes into the training to find the interconnection weights between the layers. However, the RBF is a more refined structure, and the neural network is trying to turn the training data into clusters. Generally speaking, the time required for training is faster for an RBF network than an BP network.

In the evaluation of using more input variables in the predictor, it seems that the use of more inputs is not helping to give a more accurate prediction of systolic blood pressure. Also, the database may not be large enough for the effect of the variable to be visible.

VI. Conclusions

This paper has proposed a big data intelligence analysis on biomedical application, which is for the prediction of the systolic blood pressure of a person. The work is implemented by the back-propagation (BP) neural network and the radial basis function (RBF) network. The results indicate that a blood pressure predictor can be developed based on the artificial neural network models proposed. The average prediction error (absolute difference between the predicted value and measured value) for the relationship between systolic blood pressure and the input attributes is at an acceptable level.

Our results indicate that the machine learning technique, which is a branch of big data technology, can be an efficient tool for analyzing the relationship between the systolic blood pressure and health factors (serum cholesterol, fasting blood sugar and electrocardiographic signal). It must be emphasized that the developed model does not aim to replace the direct measurement of blood pressure of a person. However, the method would provide an estimated value if the person is at a remote location. It would also provide a reference value to compare with the measured value if the measurement is taken. The method contributes to the development of a SBP predictor, which can provide early warning to hypertension and cardiovascular disease risks.

REFERENCES

[1] S.D. Yung. "Behavioral insights on big data: using social media for predicting biomedical outcomes," Trends in Microbiology, vol. 22, 2014.

[2] F.F.Costa. "Big data in biomedicine," Drug Discovery Today, vol. 19, 2014.

[3] World Health Organization, Q&As on hypertension, http://www.who.int/features/qa/82/en/

[4] Difference Between Hypertension and High Blood Pressure, http://www.differencebetween.net/science/health/difference-between-hypertension-and-high-blood-pressure.

[5] Normal Blood Pressure Range by Age, available at; http://www.ehow.com/facts_4829234_normal-blood-pressure-range-age.html.

[6] S.H. Sohan, "Management of ischemic optic neuropathies," Indian J Ophthalmol. vol.59, pp. 123–136, 2011.

[7] D. Sparrow, A.J. Garvey, B. Rosner, Jr. H.E. Thomas. "Factors in predicting blood pressure change," J Circulation. vol.65, pp.789-794, 1982.

[8] V. Florey Cdu, R.M. Acheson. "Blood pressure as it relates to physique, blood glucose, and serum cholesterol," Vital Health Stat. vol. 34, pp.1-29, 1969.

[9] I. Hjermann, A. Helgeland, I. Holme, P.G. Lund-Larsen, P. Leren. "The association between blood pressure and serum cholesterol in healthy men: the Oslo study," J Epidemiol Community Health. vol.32, pp. 117–123, 1978.

[10] C.D. Conway, N. Lavorato, F. Cunha, C. Belchior. "Modelling of the relationship between systolic blood pressure and glucose with the magnesium ion present in the blood plasma: an approach using artificial neural networks," Health. vol.1, pp. 211-219, 2009.

[11] S. Haykin. Neural Network: A Comprehensive Foundation. New Jersey: Pearson Education; 1999.

[12] S.Y. Kung, M.W. Mak, S.H. Lin. Biometric Authentication: A Machine Learning Approach. New Jersey: Prentice Hall, 2004.

[13] D. Graupe. Principles of Artificial Neural Network. World Scientific, 2013.

[14] D.X. Hao, R. Chen. "AFSs-RBF Neural Network for Predicting Earthquake-induced Liquefaction of Light Loam. Natural Computation (ICNC)," vol.3, pp. 1518-1522, 2010.

[15] M. Kaminski, T. Orlowska-Kowalska. "Clustering in optimization of RBF-based neural estimators for the drive system with elastic joint. Industrial Electronics," pp.1907 – 1912, 2011.

[16] D.X. Niu, H. Shi, J.Q. Li, Y.N. Wei. "Research on Short-term Power Load Time Series Forecasting model Based on BP Neural Network. Advanced Computer Control," pp.509-512, 2010.

[17] B.M. Wilamowski. "Suitability of fuzzy systems and neural networks for industrial applications," Optimization of Electrical and Electronic Equipment, pp.1-7, 2012.

[18] Z.R. Yang. "A Novel Radial Basis Function Neural Network for Discriminant Analysis. Neural networks," vol.17, pp. 604-612, 2006.

[19] F. Heimes, B. van Heuveln. "The Normalized Radial Basis Function Neural Network. Systems, Man, and Cybernetics," vol.2, pp. 1609-1614, 1998.

[20] UCI Machine Learning Repository, available at: http://archive.ics.uci.edu/ml/datasets/Heart+Disease.