Research

How to estimate and predict the expenses incurred by diabetes treatment using Artificial Neural Network (ANN)

Mohammad Mahboubi[1], Mehrali Rahimi[2], Mahshid Mohebbi[3], Fariba Ghahramani[4*]

[1] PhD in Health Services Administration, Kermanshah University of Medical Sciences, Kermanshah, Iran

[2] Specialist in Endocrinology and Metabolism, Kermanshah, Kermanshah University of Medical Sciences, Kermanshah, Iran

[3] MSc in Accounting, Marvdasht Islamic Azad University, Iran

[4] MSc in Epidemiology, Shiraz University of Medical Sciences, Shiraz, Iran

*correspondence should be addressed to Fariba Ghahramani, Department of Epidemiology, School of Health, 16 alleys, Alzahra Street, Shiraz, Iran; Tell: +987117251001; Fax: +987117260225; Email: ghahramani@sums.ac.ir.

ABSTRACT

Diabetes is considered a great health problem due to its economic importance and the fact that it is a chronic disease. The aim of this study was to determine the costs imposed on diabetic patients using Artificial Neural Network. The study data were gathered by randomly investigating 396 individuals who referred to Kermanshah diabetic center. In this study, Artificial Neural Network using Multiple Layer Perception (MLP) was used to investigate the costs imposed on diabetic patients. In this research, the variable related to treatment of diabetes was calculated through neural network in 8 different output layers. The eight output layers of expenses included physician's visit, medication, tests, hospitalization, radiology, treatment of symptoms, transportation, and counseling. In this study, the highest annual expenses were related to medication, tests, and radiology. In addition, the patients who presented the symptoms spent more money for treatment. Considering need to efficiently use medical facilities, it is necessary use ways which be applied to predict the medical expenses. We came to conclusion that neural network had great advantages over the regression model and could be used as an efficient tool for prediction of expenses and can replace the classical and statistical models.

Key words: Neural Network Model, Diabetes, Costs

1. INTRODUCTION

Diabetes is considered as one of the great health problem due to its economic importance and the fact that it is a chronic disease. Taking preventive measures and weakening the side effects can be helpful in reducing the expenses imposed upon the health systems (1). Studies which have been carried out in Europe and the United States suggest that the expenses imposed by diabetics are far more than those by non-diabetics (2). In low and middle income countries, less attention is paid to the economic importance of the disease, while in these areas; diabetes is one of the most important non-contagious diseases that have a negative effect on the economic growth (3). In a study carried out in Isfahan in 1999, the annual expense for the patients who injected insulin, used oral treatment, and were treated by diet was US$ 2003, US$ 76, and US$ 34, respectively. The results indicated that diabetes treatment could be how expensive. In addition, the annual expense was 1523±14.5 for each diabetic, while 52±5.8 for each subject in the control group, which was 2.92 times more in diabetics. The main expense was related to medicine, medical supplies, and hospitalization. Besides, 53% of the direct expenses were due to the diabetes side effects.

Overall, that study revealed that diabetes is a high cost medical problem in Iran (3). Considering the high prevalence of diabetes in Iran, the aim of our research was to estimate the annual expense of diabetes. Apparently, lack of systems for assessing the expenses based on health and medical information is one of the main sources of problem in evaluating medical treatment. As a result, by having an evaluative system, many of the problems will be solved. Therefore, this research is trying to find a way to estimate the medical expenses using the existing information, data processing algorithms, and accurate evaluation of the costs through Artificial Neural Network (ANN). ANN is a mathematical construct which has the ability to show the processes and non-linear combinations between the input and output of each system. This network has been designed by the data existing during the learning process and is used for future predictions (4). In this study, neural network system with multilayer perceptron was used to assess the expenses. Perceptron network is made based on a measurement unit called perceptron. One ventral perceptron takes an input with real amount and calculates a linear combination of the input. This Multilayer Perceptron (MLP) network includes an input layer, one or a few hidden layers, and an output layer. To teach this network, back propagation algorithm is used. During MLP network teaching by using back propagation algorithm, first, the calculations are done from the network input to output and then, the calculation errors are sent to the previous layers. At first, output calculation is done layer by layer and the output of each layer becomes the input of the next one. In BP, first the output layers are modified because for each neuron of the output layer, there is an optimum amount with those help and with updating rules one can modify the weights (5). Nowadays, genetic algorithm is being more commonly used to investigate the health related problems and the frequent variables. ANN is a statistical method for research and improvement. Genetic algorithm is a part of evolutional calculation which is a part of artificial intelligence. Special features of this algorithm cause it not to be considered as a random searcher. In fact, the initial idea of this method comes from Darwin's theory of evolution and its application is based on natural genetics (6). Up to now, only a limited number of researches have been conducted on this issue in Iran. Moreover, no studies have been performed on the issue in Kermanshah so for. Thus, the present study aims to design a neural network to investigate and predict the expenses caused by treatment of the patients referring to diabetic research center in Kermanshah. This research mainly aims to predict the treatment expenses with above 98% accuracy and present the evaluation model of the expenses using ANN in order to be used in treatment systems. It also aims at investigation of the effect of the variables, such as age, sex, occupation, income, symptoms, and the duration of the illness, which are related to the treatment expenses incurred upon the patients. The advantages of using ANN over using statistical methods, such as Logistic Regression, have been pointed out in various studies (7, 8). Therefore, in this study, ANN was used to predict the treatment expenses. Some of the applications of ANN in medical sciences are lowering the expenses, increasing the reliability and accuracy of physicians in their decision-makings, and designing more efficient medical supplies. ANN can predict the treatment expenses and duration. This method, as well can be applied to predict the expenses and the duration of treatment in the patients suffering from the symptoms.

2. MATERIALS AND METHODS

The present investigation was conducted on 396 subjects randomly selected from the patients referring to Kermanshah diabetes center. The annual treatment expense was chosen as the dependent variable and other variables were entered into the system as factors. In addition, the independent and dependent variables were considered as input and output layers, respectively. The output layer was the treatment expenses, while the input layers were age, sex, occupation, treatment duration, the type of treatment, and the symptoms. In this study, ANN using Multiple Layer Perception (MLP) was used to investigate the diabetes treatment. All the statistical analyses were performed through the SPSS statistical software (version 20). The process of the neural network was fitted for the dependent variable; i.e., treatment expenses, and other variables were calculated as factors. To investigate the impact of the independent variables on the total treatment expense, the classic statistical tests were initially used. In these tests, many errors can occur, while the ANN investigation of the relationship between the variables does not show any errors. Also, the prevention of data defect is highly possible in ANN (9). This method is based on mathematical relationships between the data and aims at investigation of a large number of independent and dependent variables as well as their impact

on each other (10). In this study, multilayer perception ANN with BP was used for prediction. This network can accurately estimate any relationship between the input and output. In this study, the best conformity model which could calculate with the least number of errors and the best approximation was determined. The model has to be assessed regarding the output results and by assessment of the model, we aim to determine how capable it is to give a reliable result related to the input. The current research made use of absolute error percentage index which has also been chosen by many researchers to evaluate the prediction. In this study, the number of samples was determined in each output data, valid data, and test and learning in neural network processes (Table 1).

Table1. Summary of neural network processes

	Number	Percent
training	287	72.5
Testing	109	27.5
Valid	390	
Excluded	6	
total	396	

In this memory or neural network, nodes have two states: active (on or 1) and inactive (off or 0). Besides, each synopsis on the relationship between the nodes has the same weight. The synapses with positive weight stimulate or activate the next inactive node, while the synapses with negative weight inactivate or control the next joined node (in case it has been active).

3. RESULTS AND DISCUSSION

In this study, the data related to 496 diabetics who had referred to Kermanshah diabetic research center were investigated. According to the results, 65.9% of the participants were female and 34.1% were male. The mean age of these patients was 53.9±11.7 years. In addition, the mean of the subjects' disease duration was 9±6.1 years. In this research, neural network was used to calculate the variables related to treatment of diabetes in 8 different output layers in the form of parameters, such as age, sex, occupation, duration of the disease, symptoms, type of insurance, and type of treatment. To evaluate the neural network results, r=0.45 and r=0.7 were predicted for age and duration of the disease, respectively using linear regression and these two variables were entered into the

model as the input layer. By using neural network, the importance of the independent variables such as duration of the disease and the side effects which had the highest impact on the expenses was determined (Table 2).

Table 2. The importance of the independent variables On the cost of Diabetes

Variable	Normed importance (%)	Important factor
Sex	23.1	**0.127**
Complication	36.3	**0.199**
Age	23.1	**0.126**
Duration of disease	100	**0.548**

Eight output layers of the expenses included physician's visit, medication, tests, hospitalization, radiology, treatment of symptoms, transportation, and counseling. In this study, the highest annual expenses were related to medication, tests, and radiology (2514369, 1010447, and 971282 US dollars, respectively). On the other hand, the lowest expense was 15666 US dollars and related to counseling. The patients who showed the symptoms spent more money for the treatment. The most common symptoms were Macroscolar, Microscolar, and cataract. The results indicated that 21.5% of the patients either used insulin alone or along with pills and the rest used pills alone. The average expense of the treatment with insulin was 1.5 times more than that with pills. The population of Kermanshah is 851405 according to the last national census which took place in 2011. In the present research, the prevalence of diabetes was 9.2% in this city (11). Moreover, this study showed that the annual expense for diabetic patients was 160 million Rials (more than 6 million US dollars) in Kermanshah. Considering the average monthly income of the families and each individual's expenses, 4% of each family's income is spent on diabetes treatment. To obtain the best MLP neural network, different models were created, tested, and taught. In the end, two-layer network with 4 neurons in the first layer (hidden layer), 3 neurons in the second layer (hidden layer), and a neuron in the third layer (output layer) was introduced as the best network model for prediction (figure1).
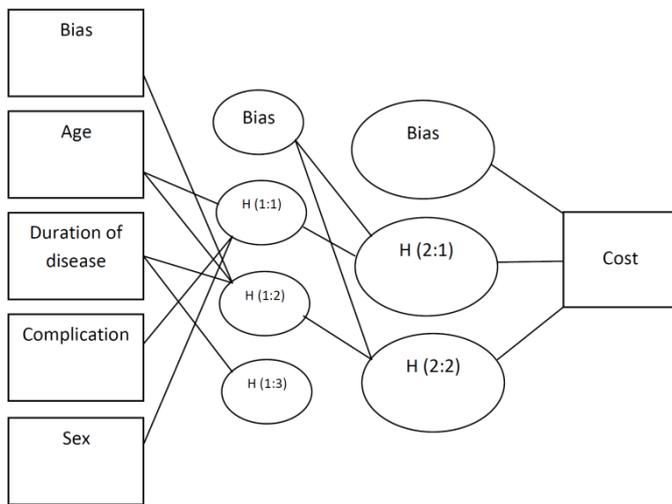
Figure1. Neural network architecture for the cost of treating diabetes

In early 1980s, researches were mainly focused on the estimability of the data. Up to now, a large number of studies have been conducted on the healthcare data in order to predict the expenses. In 2000, Van de van presented the risk-based expense prediction model. At the moment, various models are available to insurance companies to predict the expenses. These models are extracted from the data related to disease diagnosis and doctors' prescription. Moreover, the error rate of the data processing algorithms is less than that of regression and analysis of variance (12). Up to now, numerous researches have been done to assess the hospitalization expenses by using regression and variance analysis. Although this research was based on only numbers of available variables and very few variables play roles in hospital expenses, the accuracy of these assessments was less than expected (13). In this research, the total expenses of diabetes were simulated using neural network. The results showed that neural networks could be used as a powerful tool for modeling and estimating the treatment expenditures. The results also showed that the duration of the disease, age, and symptoms had a great impact on the treatment expenses. Since the expenses were investigated as 8 output layers, 8 neural networks were designed. Considering the results of this research, it was concluded that the treatment expense of a diabetic person in Kermanshah is quite high; in a way that both the patients and insurance companies have to spend a lot of money on the disease annually. By controlling the symptoms of the disease, one can lower the expenses to a great extent. For instance, research has shown that 33.386 US dollars were saved in the United States (14). In the present study, the annual expense was 82 US dollars for each individual. Yet, it was reported as 152.3±14.5 dollars in

another research which was conducted in Tehran in 2006 (3). This difference was due to the difference in dollar exchange rate in 2006. In a study carried out in the U.S. in 2007, the annual expense of each diabetic patient was 700 dollars (15). The difference in the expenses can be due to the dollar exchange rate in different countries. One study which was conducted in Pakistan showed that the main expenses of diabetes treatment were related to medicine (46%) followed by lab tests (32%). This finding as well as the result related to age and disease duration was consistent with the present research (16). Furthermore, a study in Norway in 2005 showed that the main expenses of diabetes treatment were due to medication and medical supplies (17). In 2004, a study on the difference between the categorization techniques and neural networks was conducted on 492 patients in order to assess the patients' expenses. The variables included hospitalization, the number of surgeries, age, sex, the physician's name, the number of referrals to a counselor, and the number of radiologies. After modeling, it was concluded that neural network algorithm was better than the categorization algorithm (13). In 2009, a more detailed research was done to estimate the expenses incurred on cancer patients. It was revealed that in neural network algorithm, the expenses were estimated with 98% accuracy. The most important neural network variables for predicting the expenses were hospitalization period and the number of medical counseling and group therapies. On the other hand, the most important variables for the categorization method were hospitalization period, the number of referrals to ICU, and the number of medical counseling (18). In addition to application of data analysis for predicting the treatment expenses, data analysis was also used to predict heart attack and to allocate medical care resources for the patients suffering from lung cancer (19-21). Considering the need to efficiently use the medical facilities, it is also necessary that some ways be used to predict the medical expenses. In fact, there will be fewer errors, damages, and casualties if the medical sciences are equipped with smart tools to diagnose and cure diseases. Since neural network is capable of non-linear modeling, it can well predict the sudden changes. It may also help modeling and prediction to a great extent (22).

4. CONCLUSION

In the end, we came to the conclusion that neural network has great advantages over the regression model, and can be used

as an efficient tool for prediction of the expenses and can replace the classical and statistical models. Moreover, prevention and control of the disease can tremendously lower the radiology, test, medication, and the total expenses of the treatment.

ACKNOWLEDGMENT

AUTHORS CONTRIBUTION

This work was carried out in collaboration between all authors.

CONFLICT OF INTEREST

Authors have declared that no conflict interests exist.

REFERENCES

1. Shahrad Bejestani H, Motabar A. Assessment of Diabetic Foot Ulcer's Predisposing Factors and its Outcomes in Patients with Diabetic Foot Syndrome Hospitalized in Hazrat Rasoul-e-Akram Hospital in Tehran During 1996-2001. Razi Journal of Medical Sciences. 2004;11(39):77-83.

2. Köster I, Von Ferber L, Ihle P, Schubert I, Hauner H. The cost burden of diabetes mellitus: the evidence from Germany—the CoDiM Study. Diabetologia. 2006;49(7):1498-504.

3. Esteghamati A, Khalilzadeh O, Anvari M, Meysamie A, Abbasi M, Forouzanfar M, et al. The economic costs of diabetes: a population-based study in Tehran, Iran. Diabetologia. 2009;52(8):1520-7.

4. Menhaj M. Fundamentals of Artificial Neural networks. Vol.(1). Amir Kabir University Publishing. Tehran, Iran. 1998.

5. Chayjan RA, Montazer G, Hashjin TT, Khoshtaghaza M, Ghobadian B. Prediction of pistachio thermal conductivity using artificial neural network approach. International Journal of Agriculture & Biology. 2007;9:816-20.

6. Deb K. An introduction to genetic algorithms. Sadhana. 1999;24(4-5):293-315.

7. Sedehi M, Mehrabi Y, Kazemnejad A, Hadaegh F. Comparison of Artificial Neural Network, Logistic Regression and Discriminant Analysis Methods in Prediction of Metabolic Syndrome. Iranian Journal of Endocrinology & Metabolism. 2010;11(6).

8. Recuenco S, Eidson M, Cherry B, Johnson G. Risk-Based Cost Modelling of Oral Rabies Vaccine Interventions for Raccoon Rabies. Zoonoses and public health. 2009;56(1):16-23.

9. Kasabov NK. Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering, A Bradford Book. The MIT Press, Cambridge, Massachusetts, London, England; 1998.

10. Stern HS. Neural networks in applied statistics. Technometrics. 1996;38(3):205-14.

11. Lotfi MH, Saadati H, Afzali M. Prevalence of Diabetes in People aged≥ 30 years: The Results of Screen-ing Program of Yazd Province, Iran, in 2012. Journal of research in health sciences. 2013;14(1).

12. Bertsimas D, Bjarnadóttir MV, Kane MA, Kryder JC, Pandey R, Vempala S, et al. Algorithmic prediction of healthcare costs. Operations Research. 2008;56(6):1382-92.

13. Lee S-M, Kang J-O, Suh Y-M. Comparison of hospital charge prediction models for colorectal cancer patients: neural network vs. decision tree models. Journal of Korean medical science. 2004;19(5):677-81.

14. Huang ES, Zhang Q, Brown SE, Drum ML, Meltzer DO, Chin MH. The Cost-Effectiveness of Improving Diabetes Care in US Federally Qualified Community Health Centers. Health services research. 2007;42(6p1):2174-93.

15. Dall TM, Zhang Y, Chen YJ, Quick WW, Yang WG, Fogli J. The economic burden of diabetes. Health Affairs. 2010;29(2):297-303.

16. Khowaja LA, Khuwaja AK, Cosgrove P. Cost of diabetes care in out-patient clinics of Karachi, Pakistan. BMC Health Services Research. 2007;7(1):189.

17. Solli O, Jenssen T, Kristiansen IS. Diabetes: cost of illness in Norway. BMC endocrine disorders. 2010;10(1):15.

18. Kang JO, Chung S-H, Suh Y-M. Prediction of hospital charges for the cancer patients with data mining techniques. Journal of Korean Society of Medical Informatics. 2009;15(1):13-23.

19. Srinivas K, Rani BK, Govrdha A. Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. International Journal on Computer Science & Engineering. 2010;2(2).

20. Patil SB, Kumaraswamy Y. Intelligent and effective heart attack prediction system using data mining and artificial neural network. European Journal of Scientific Research. 2009;31(4):642-56.

21. Phillips-Wren G, Sharkey P, Dy SM. Mining lung cancer patient data to assess healthcare resource utilization. Expert Systems with Applications. 2008;35(4):1611-9.

22. Ahmadi M, Farhadi B, Nazifi M, editors. Evaluation the forecasting of ANN method in estimation evalaporatation and transpiration. Proceedings of the forth conference of water management; 2011.