www.banglajol.info/index.php/JSSMC

Original Article

J Shaheed Suhrawardy Med Coll, December 2014, Vol.6, No.2 ISSN 2226-5368

Comparison of Predictive Models in Classification of Nutritional Status among Arsenic Affected People in Rural Areas in Bangladesh: An Artificial Neural Network (ANN) and Logistic Regression Approach

Rahman A[1], Akter M[2], Majumder AK[3], Haque MA[4], Sattar AFMA[5]

Abstract

Background: Clinical data play an important role in medical sector for binary outcome variables. Various methods can be applied to build predictive models for the clinical data with binary outcome variables. **Objective:** This research was aimed to explore and compare the process of constructing common predictive models. Methodology: Models based on an artificial neural network (the connectionist approach) and binary logistic regressions were compared in their ability to classifying malnourished subjects and those with over-weighted participants in rural areas of Bangladesh. Subjects were classified according to the indicator of nutritional status measured by body mass index (BMI). This study also investigated the effects of different factors on the BMI level of adults of six villages in Bangladesh. Demographic, anthropometric and clinical data were collected based on aged over 30 years from six villages in Bangladesh that were identified as mainly dependent on wells contaminated with arsenic. **Result:** A total of 460 participants were recruited for this study. Out of 460(140 male and 320 females) participants 186(40.44%) were identified as malnourished (BMI<18.5 gm), and the remainder 274(59.56%) were found as over-weighted (BMI>18.5 gm). Among other factors, arsenic exposures were found as significant risk factors for low body mass index (BMI) with a higher value of odds ratio. This study shows that, binary logistic regression correctly classified 72.85% of cases with malnourished in the training datasets, 76.08% in the testing datasets and 75.26% of all subjects. The sensitivities of the neural network architecture for the training and testing datasets and for all subjects were 84.28%, 84.78% and 81.72% respectively, indicate better performance than binary logistic regression model. **Conclusion:** This study demonstrates a significant performance of artificial neural network than the binary logistic regression models in classification of malnourished participants from over-weighted ones. (J Shaheed Suhrawardy Med Coll, 2014;6(2):71-75)

Keywords: Artificial Neural Network (ANN), binary logistic regression, classification, malnourished, over-weighted

Received: December 2013; Revised: March 2014; Accepted: November 2014

Introduction

Brain cells produce tiny electrical signals known as impulses. Artificial Neural Network (ANN) modeling is a paradigm for computation and knowledge representation. It is originally inspired by the aspect of the information processing and physical structure of the brain with a web of neural connection. Therefore some writers classified it as a "microscopic", "whole box" system and an expert system as a "microscopic", "black-box" system[1]. In Bangladesh, a population of some 30-70 million people living in 41 districts out of the 64 are probably exposed to arsenic from drinking water containing >50mg/L arsenic for a long period[2]. The exposure probably started in late 1960s when drilling of tube wells began as part of a wide irrigation plan[3-6]. In another study, Rahman further examined the relation between arsenic exposure and glucosuria taken as a proxy for diabetes mellitus in subjects. Lower body mass index (BMI) was reported among the arsenicosis patients compared to the

1. Azizur Rahman, Lecturer, Department of Statistics, Jahangirnagar University, Dhaka, Bangladesh
2. Mariam Akter, Teaching Assistant, School of Business, United International University, Dhaka, Bangladesh; Email: mariam.rahman576@gmail.com
3. Prof. Ajit Kumar Majumder, Professor, Department of Statistics, Jahangirnagar University, Savar, Bangladesh; Email: ajitm@ju.ac.bd
4. Dr. Md. Atiqul Islam, Assistant professor, Department of Statistics, Banghabandhu Shiekh Mujib Medical University, Dhaka
5. Dr.AFM Arshedi Sattar, Assistant Professor, Department of Microbiology, National Institute of Neurosciences & Hospital, Dhaka

Correspondence
Azizur Rahman, Lecturer, Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh; Email: rahman.aziz83@gmail.com;
Cell no.: +8801676718764

Conflict of interest: None
Financial Support: This research project was not funded by any group or any institution. There is no external funding.

unexposed population in Bangladesh[4].

Thus there is an urgent need to identify and manage patients in rural areas with arsenic exposure having poor nutritional status, especially in groups at higher risk for arsenic related health effects and its complications[5]. Therefore the purpose of the present study was to explore and compare the process of constructing common predictive models.

Methodology

The study was cross-sectional and was performed by the door-to -door visits to interview families with known arsenic concentration in their wells. Six villages in two districts of Bangladesh named Comilla and Jhenidah were selected for the study on the basis of existing survey reports of arsenic measurements in drinking water. Those who had lived in the study areas throughout their lifetimes and who had used the same well as long as it had existed were selected as eligible subjects. Among this population, aged>30 years who had record of BMI <18.5 gm and had complete information were the subjects of the presents study. Data were collected after obtaining the necessary approval from the villagers; participants were informed about the objectives before the commencement of measurements. Information of age, gender, weight and height were collected on a pre-tested questionnaire. Height and weight measurements were taken on each subject following the standard techniques[6]. Individual who had the histories of arsenic exposure, were further interviewed by questionnaire and were examined for identification of nutritional status. Nutritional status was evaluated using internationally accepted BMI guidelines[5]. Body Mass Index (BMI) is used as a measure of nutritional status of each participant. Body Mass Index (BMI) was interpreted as malnourished when <18.5gm, normal if ranging between 18.5 and 24.99 gm and overweighed if it exceeds 24.99 gm. The demographic and clinical data used as predictors in the models were patients' age, sex, body mass index (BMI), number of household, history of Arsenic exposure. Arsenic exposure was defined as any prior diagnosis of this disease by a physician. The neural network development software used in this study was R, version 2.5.1 package (nnet version 7.2-290). Other statistical analyses were performed by the SPSS version 13.0. Two different models were applied to the patient data. The first was a standard binary logistic regression analysis. The second was a standard feed-forward error back-propagation multilayer perceptron with a three layer topology (input, hidden and output layers) with four neurons in the hidden layer determined by trial and error process) and no direct connection from the input to output layers[7]. The error back propagation learning algorithm is a powerful approach and despite its slow convergence, is one of the most popular and successful algorithm for pattern recognition. The two different models were compared in their ability to predict nutritional status from the participants' demographic and clinical data. The database was splited into two groups named as training data-set containing approximately 75% of the sample and testing data-set containing 25% of the subjects. Training dataset was used to develop the logistic regression and perceptron models by introducing the disease status of the subjects into the models. Testing data set was used by the models for classifying the nutritional status of subjects.

Results

A total of 460 subjects above 30 years of age were identified of which 140 (33.8%) subjects were males and 320(66.2%) subjects were females. Thus participants with a BMI<18.5, aged over 30 years, have been drinking water from a tube well and have been a resident of the study area were selected as cases. On the other hand, participants with a BMI >18.5 were recruited in the study as controls. Among 460 participants 186(40.44%) were identified as malnourished (BMI<18.5 gm) and the remainder 274(59.56%) were found as overweighed (BMI>18.5 gm). The mean age (±SD) in this study was 45.3±13.045 years overall and 47.77±14.55 years for the malnourished group (Table 1). One way ANOVA indicated that the mean age of the three groups was significantly different and Tukey post hoc multiple comparison test showed that the malnourished group was older than overweighed group.

Table 1a: Characteristics of subjects in different glucose status groups (Mean±SD)

Variables	Malnourished (n=186)	Over-weighed (n=274)	Total (n=460)
Age(in years)	47.77(±14.55)	43.62(±11.64)	45.31(±13.04)
BMI(kg/m3)	16.56(±1.38)	22.23(±3.37)	19.94(±4.11)

Table 1b: Characteristics of subjects in different glucose status groups

Sex	Malnourished (n=186)	Over-weighed (n=274)	Total (n=460)
Male	58(31.18%)	82(29.92%)	140(30.43%)
Female	128(68.81%)	192(70.07%)	320(69.57%)
Hist. of Arsenic			
Yes	160(86.04%)	147(53.64%)	307(66.73%)
No	26(13.97%)	127(46.36%)	153(33.27%)

*SD=Standard deviation, BMI= Body mass index, History of Arsenic exposure.

Those in the overweighed group had a higher mean BMI than those in the malnourished groups. The chi-squared test indicated that there was a significant association between nutritional status and history of Arsenic exposure (P<0.001).

Table 2: Distribution of nutritional status of the sample in the training and testing data sets

Variable	Training datasets	Testing datasets	Total
Malnourished	140(75.0%)	46(25.0%)	186(100.0%)
Over weighted	206(75.0%)	68(25.0%)	274(100.0%)
Total	**346(75.0%)**	**114(25.0%)**	**460(100.0%)**

*Malnourished= BMI<18.5; Over weighted= BMI>18.5

Moreover, malnourished group had a higher proportion of subjects with a positive history of Arsenic compared with the over-weighted group (86.02%, and 53.64%) for the case and control groups respectively (Table 1).

Table 3: Odds ratio and coefficients of binary logistic regression analysis of factors associated with glucose status

Characteristics	Coefficient	S.E.	OR	95.0% CI	
sex(1)	.060	.231	1.062	.675	1.672
age	-.031	.008	.969	.953	.985
HouHmem	.120	.049	1.128	1.025	1.240
parsc(1)	.507	.236	1.660	1.045	2.636
Constant	2.539	.495	12.672		

*Sex(1) and parsc(1) are categorical variables

As a common statistical method, binary logistic regression was used and indicated that all factors were significantly associated with nutritional status. Age, sex, number of house member and Arsenic exposure were significant risk factors for describing nutritional status. Those who were suffering from arsenic disease had a higher risk of malnourished (Table 3).

Table 4: Number of correct diagnosis of nutritional status using binary logistic regression model

True Status	Predicted Status using logistic-regression		
	Malnourished	**Over-weighted**	**Total**
Training Data			
Malnourished	102	38	140
Over-weighted	52	154	206
Total	**154**	**192**	**346**
Testing Data			
Malnourished	35	11	46
Over-weighted	17	51	68
Total	**52**	**62**	**114**
Overall			
Malnourished	140	46	186
Over-weighted	63	211	274
Total	**203**	**257**	**460**

Binary logistic regression correctly classified 72.85% of cases with malnourished in the training datasets, 76.08% in the testing datasets and 75.26% of all subjects (Table 4). The sensitivities of the neural network architecture for the training and testing datasets and for all subjects were 84.28%, 84.78% and 81.72% respectively (Table 5).

Discussion

Artificial neural network are used in three main ways which are (i) as models of biological nervous system and intelligence, (ii) as real-time adaptive signal processors controllers implemented in hardware for applications such as robots and (iii) as data analytic methods2. Artificial intelligence has been proposed as a reasoning tool to support clinical decision-making since the earliest days of computing[3-7]. Artificial neural networks are computer modeling technique based on the observed behaviors of biological neurons[8]. This is a non-parametric pattern recognition method which can recognize hidden patterns between independent and dependent variables[9].

Table 5: Number of correct diagnosis of nutritional status using Artificial Neural Network Architecture

True Status	Predicted Status using ANN architecture		
	Malnourished	Over-weighted	Total
Training Data			
Malnourished	118	22	140
Over-weighted	34	172	206
Total	**152**	**194**	**346**
Testing Data			
Malnourished	39	7	46
Over-weighted	10	58	68
Total	**49**	**65**	**114**
Overall			
Malnourished	152	34	186
Over-weighted	51	223	274
Total	**203**	**257**	**460**

The main principle of neural network computing is the decomposition of the input-output relationship into a series of linearly separable steps using hidden layers[6]. There are three distinct steps in developing an ANN based solution which were i) data transformation or scaling, ii) network architecture definition, when the number of hidden layers, the no of nodes in each layer and the connectivity between the nodes and set, iii) construction of learning algorithm in order to train the network[5,8]. The simple architecture of a typical network consists of an input layer, series of hidden layers, an output layer and connection between them. Notes in the input layer represent possible influential factors that affect the network outputs and have no computation activities, while the output layer contains one or more nodes that produce the network output. Hidden layers may contain a large number of hidden processing nodes. A feed forward back-propagation network propagates the information from the input layer to the output layers, compares the network outputs with known targets and propagates the error term from the output layer back to the input layer, using a learning mechanism to adjust the weights and biases[5,10].

In 1957, Rosenblatt invented the perceptron, an artificial neuron, in which dendrites are replaced by weighted inputs that are summed inside the artificial neuron and pass through a suitable threshold (activation)[10]. The activated outputs transfer from inner to output layers and produce an output to simulate a desired output (target) at the end. By a learning algorithm, the neural net achieves a form of learning by modifying weights proportional to the difference between the target and the gained output[11]. Artificial neural network have been applied to diagnosis and decision-making in various medical fields[12-16]. Statistical methods such as discriminate analysis and logistic regression have commonly been used to develop models for clinical diagnosis and treatment5. But studies published in recent years have reported that the artificial neural networks approach improves prediction in several situations including prognosis of breast cancer in women after surgery[17], modeling for surgical decision-making for patients with traumatic brain injury[5] and survival of alcoholic patients with severe liver disease[16]. In contrast, others have reported that artificial neural networks and statistical models yielded similar results[9,18].

World Health Organization has recommended that anthropometry could be used to assess the nutritional and health status of adults. One such measure now in widespread use in Quetelet's index, which is body weight (kg) divided by stature (m2)[19]. Better known as body mass index (BMI), this measure was an attempt by the 19th century mathematician Lambert Adolphe Jacques Quetelet to describe the relation between body weight and stature in humans[20]. Many studies have shown that BMI is reasonable measure of adiposity[21-24]. A low BMI and high level of under nutrition (based on BMI) is a major public health problem especially among rural underprivileged adults of developing countries[25]. Although adult nutritional status can be evaluated in many ways, the BMI is most widely used because its use is simple, inexpensive, safe and suitable for large scale surveys[26-30]. Thus, BMI is the most established anthropometric indicator used for assessment of adult nutrition status[31].

Limited studies have indicated that poor nutritional status may increase the risk of arsenic related health effects[32-35]. Participants with poor nutritional status weight below 80% of the standard body weight for their age and sex were reported form West Bengal, India to have an overall 1.6 fold increase for male=1.5, females=2.1 in the prevalence of keratoses, suggesting that malnutrition may increase the susceptibility for arsenic toxicity[36]. Arsenic affected people of south western Taiwan and the Antofasta region in northern Chile were reported to have a low socio-economic status and poor nutritional status[37-40]. Lower Body mass index (BMI) was reported among the arsenicosis patients compared to the unexposed population in a previous study of Bangladesh[41]. Here, a significant trend for increased risk of malnourished was observed for increasing dosage of arsenic exposure and the subjects were more accurately identified with the help of ANN approach rather than linear regression model. In this study, the primary database of the patients was used to develop models to try to distinguish subjects with malnourished from over-weighted subjects. The accuracy of the perceptron and binary logistic regression models in predicting a subject's glucose status were compared. Here, binary logistic regression correctly classified 72.85% of cases with malnourished in the training datasets, 76.08% in the testing datasets and 75.26% of all subjects. The sensitivities of the neural network architecture for the training and testing datasets and for all subjects were 84.28%, 84.78% and 81.72% respectively.

Conclusion

In the conclusion this study demonstrates a significant performance of artificial neural network than the binary logistic regression models in classification of malnourished participants from over-weighted ones.

References

1. Eldon YL. Artificial Neural Networks and Their Business Applications. Inf Mang 1994, 27(5): 303- 313
2. Warren SS. Neural Networks and Statistical Models. Proceedings of the 19the Annual SAS Users Group International conference, April 1994, USA
3. Alonso-Betanzos A et al. Applying statistical uncertainty-based and connectionist approaches to the prediction of fetal outcome: a comparative study. Artificial intelligence in medicine, 1999;17(1):37-57
4. Lisboa PJA. A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Networks 2002;15:11-39
5. Li YC, Chiu WT, Jian WS. Nueral networks modeling for surgical decisions on traumatic brain injury patients. Internat J Med Informat 2000;57:389-405
6. Lohman TO, RocheAF, Martorell R. Anthropometric Standardization Reference Manual. Chicago: Human Kinetics Books, 1988
7. Bishop CM. Neural networks for pattern recognition. Oxford, Oxford University Press, 1995
8. Kazemnejad A, Batvandi Z, Faradmal J. Comparison of artificial neural network and binary logistic regression for determination of impaired glucose tolerance/diabetes. Eastern Mediterranean Health J 2010;16(6)
9. Ergun U et al. Classification of carotid artery stenosis of patients with diabetes by neural networks and logistic regression. Computers Biology Med 2004;34: 389-405
10. Rosenblatt F. The perceptron: a perceiving and recognizing automation. Cornell Aeronautical Laboratory report 85-460-I. Ithaca, New York, Cornell Aeronautical Laboratory, 1957
11. Bishop CM. Neural networks for pattern recognition. Oxford, Oxford University Press, 1995
12. Ronco AL. Use of artificial neural networks in modeling associations of discriminant factors: towards an intelligent selective breast cancer screening. Artificial Intelligence Med 1999;16(30): 299-309
13. Kennedy RL. An artificial neural network system for diagnosis of AMI in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Computer Methods Programs Biomed 1997;52(2): 93-103
14. Cross SS. Image analysis of low magnification images of fine needle aspirates of the breast produces useful discrimination between benign and malignant cases. Cytopathology 1997;8: 265-73
15. Dybowski R, Gant V. Artificial neural network in pathology and medical laboratories. Lancet 1995;346:1203-7
16. Lapuerta P, Rajan S, Bonacini M. Neural networks of outcomes in alcoholic patients with severe liver disease. Hepathology 1997;25: 302-6
17. Lisboa PJA. A Bayesian neural networks approach for modeling censored data with an application to prognosis after surgery for breast cancer. Artificial Intelligence Med 2003;28(1):1-25
18. Tafeit E. The determination of three subcutaneous adipose tissue compartments in non-insulin-dependent diabetes mellitus women with artificial neural networks and factor analysis. Artificial Intelligence Med 1999;17: 181-93
19. Keys A, Fidanza F, Karvonen MJ et al. Indices of relative weight and obesity. J Chronic Dis 1972; 25:329-43
20. Quetelet LAJ. A treatise on man and the development of his faculties. In: Comparative statistics in the 19th century. Ediburgh: William and Robert Chambers,1842
21. DeFronzo RA, Bonadonna RC, Ferrannini E. Pathogenesis of NIDDM. In: Alberti KGMM, Zimmet P, DeFronzo RA, Keen H. (Eds.), International Test book of Diabetes Mellitus, second ed. Wiley, New York, 1997; 635-711

22. Tseng C H, Tseng CP, Chiou HY, Hsueh YM, Chong CK, Chen CJ. Epidemiologic evidence of diabetogenic effect of arsenic, Toxicology letters 2002;133:69-76
23. Rahman M, Tondel M, ahmed SA, Azelson O. Diabetes mellitus associated with arsenic exposure in Bangladesh, American J Epidem 1996;148:196-203
24. Kazemnejad A, Batvandi Z, Faradmal J. Comparison of artificial neural network and binary logistics regression for determination of impaired glucose tolerance/diabetes. Eastern Mediterranean Health Journal 2010;16(6)
25. World Health Organization. (1995). Physical Status: the Use and Interpretation of Anthropometry. Technical Report Series no. 854. Geneva: World Health Organization
26. Bose K. Generalized Obesity and Regional Adiposity in Adult White and Migrant Muslim Males from Pakistan in Peterborough. J Royal Society Promotion Health 1996;116:161-167
27. Deurenberg P, Weststrate AJ, Seidell CJ. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. British J Nutrit 1991;65:105-114
28. Ferro-Luzzi A, Sette S, Franklin M, James TPW. A simplified approach of assessing adult chronic deficiency. European J Clin Nutrit 1992;46:173-186
29. Garrow JS, Webster DJ. Quetelets's index (W/H) as a measure of fatness. International J Obesity 1995;9:147-153
30. James TPW, Mascie-Taylor NGC, Norgan NG, Bristrian BR, Shetty PP, Ferro-Luzzi A. The value of arm circumference measurements in assessing chronic energy deficiency in Third World adults. European J Clin Nutrit 1994;48:883-894
31. Khosla T, Lowe R. Indices of obesity derived from body weight and height. Br J Prev Soc Med 1967;21:122-128
32. Lee RD, Nieman, DC. Nutritional Assessment. 2nd Edition, New York: McGraw Hill, 2003
33. Lohman TO, Roche AF, Martorell R. Anthropometric Standardization Reference Manual. Chicago: Human Kinetics Books, 1988
34. Smalley KJ, Knerr AN, Kendrick ZV, et al. Re-assessment of body mass indices. American J Clin Nutrit 1990;52:405-408.
35. Strain GW, Zumoff B. The relation of weight-height indices of obesity to body fat content. J American Coll Nutrit1991;11:715-718
36. Das S, Bose K. Body mass index and chronic energy deficiency among adult Santals of Purulia District, West Bengal, India. Internat J Human Sci 2010;2(2):488-503
37. Abul HM, Shahidullah SM, Smith W, Hossain KS, Hasan Z, Ahmed KT. Association between chronic Arsenic exposure and nutritional status among the women of child bearing age: A case-control study in Bangladesh. Internat J Environmen Res Public Health 2010;7:2811-2821.
38. Abul HM, Shahidullah SM, Smith W, Hossain KS, Hasan Z, Ahmed KT. Association between nutritional status and arsenicosis due to chronic arsenic exposure in Bangladesh. Internat J Environmen Res Public Health 2004;14, 99-108
39. Shetty PS. Body Mass Index, a measurement of Chronic Energy Deficiency in Adults; FAO: Rome, Italy, 1994
40. Borgono JM, Vicent P, Venturino H, Infante A. Arsenic in the drinking water of the city of Antofagsta: Epidemiological and clinical study before and after the installation of a treatment plant. Environmental Health Perspective 1977;19:103-105
41. Hsueh YM, Cheng GS, Wu MM, Yu HS, Kuo TL, Chen CJ. Multiple risk factors associated with arsenic-induced skin cancer: effects of chronic liver disease and malnutrition status. Brazilian J Cancer 1995;71:109-114