ORIGINAL ARTICLE

Application of Connectionist Approach in Classification of Nutritional Status Among Arsenic Affected people in Rural Areas in Bangladesh

*A Rahman[1], M Akter[2], A K Majumder[3]

[1]*Azizur Rahman, Lecturer, Department of Statistics, Jagannath University, Dhaka-1100, Bangladesh*
[2]*Mariam Akter, Teaching Assistant, School of Business, United International University, Dhaka Bangladesh*
[3]*Ajit Kumar Majumder, Professor, Department of Statistics, Jahangirnagar University, Savar Bangaldesh*

*Corresponding Author

ABSTRACT

Various methods can be applied to build predictive models for the clinical data with binary outcome variables. This research aims to explore and compare the process of constructing common predictive models. 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 a sample population of 460 adults of six villages in Bangladesh. Demographic, enthropometric and clinical data were collected based on a total of 460 participants aged over 30 years from six villages in Bangladesh that were identified as mainly dependent on wells contaminated with arsenic. 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.

Key Words: Artificial Neural Network (ANN), Binary Logistic regression, classification, malnourished, over-weighted

Introduction

Artificial Neural Network (ANN) modeling, a paradigm for computation and knowledge representation, is originally inspired by the aspect of the information processing and physical structure of the brain with a web of neural connection (see figure 1). Therefore some writers classified it as a "microscopic", "whole box" system and an expert system as a "microscopic", "black-box" system1. Artificial neural network are used in three main ways: (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, (iii) as data analytic methods[2]. 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 a computer modeling technique based on the observed behaviours of biological neurons[8]. This is a non-parametric pattern recognition method which can recognize hidden patterns between independent and dependent variables[9].

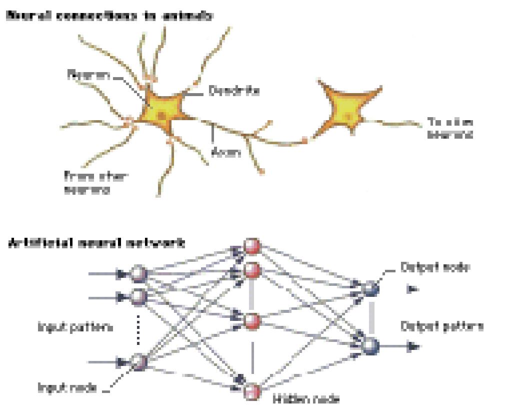

Figure-1: The Neural connection in animals (biological neuron in top) and the counterpart. Artificial Neural network structure (in bottom).

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: 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]. Figure 2 shows the simple architecture of a typical network that 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].

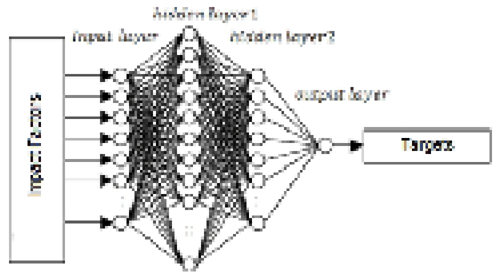

Figure-II: Simple structure of a typical neural network

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 thereshold (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 treatment[5]. 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 alchoholic 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 (1995) 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 (in kg) divided by stature (in m^2)[31]. 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[35]. Many studies have shown that BMI is reasonable measure of adiposity[26,31,37]. A low BMI and high level of undernutrition (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[28,33,34]. Thus, BMI is the most established anthropometric indicator used for assessment of adult nutrition status[33].

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[19]. The exposure probably started in late 1960s when drilling of tubewells began as part of a wide irrigation plan20. In another study, Rahman[19] further examined the relation between arsenic exposure and glucosuria (taken as a proxy for diabetes mellitus) in subjects. A recent study shows that lower body mass index (BMI) was reported among the arsenicosis patients compared to the unexposed population in Bangladesh[40]. There is thus 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[24].

Methods and Materials

Study Population

Six villages in two districts of Bangaldesh (Cumilla, Jhenidah) were selected for the study on the basis of existing survey reports of arsenic measurements in drinking water. The study was cross-sectional and was performed by the door-to -door visits to interview families with known arsenic concentration in their wells. Eligible subjects included those who had lived in the study areas throughout their lifetimes and who had used the same well as long as it had existed. A total of 460(140 males and 320 females) subjects above 30 years of age were identified. 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[34]. A total of 307 individuals had histories of arsenic exposure, were further interviewed by questionnaire, and were examined for identification of nutritional status, according to internationally accepted BMI guidelines[25]. Among this population, aged > = 30 years (33.83% male and 66.17% female) who had record of BMI < 18.5 gm and had complete information were the subjects of the presents study.

Participants' demographic and clinical characteristics

Body Mass Index (BMI) is used as a measure of nutritional status of each participant. The BMI was computed using the following standard equation: BMI = Weight(kg)/height(m^2). Nutritional status was evaluated using internationally accepted BMI guidelines[25]. The following cutoff points were used to classify the nutritional condition among subjects: malnourished (BMI < 18.5) and Over-weighted (BMI > 18.5). 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.

Prediction Models

We applied two different models 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[11]. 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[24].

The two different models were compared in their ability to predict nutritional status from the participants' demographic and clinical data. We split the database into two groups: a 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.

Software

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.

Results and Discussion

Initially a baseline survey was carried out among 460 participants aged 30 years or over, who are eligible for study purpose. Body Mass Index (BMI) was interpreted as malnourished when <18.5gm, normal if ranging between 18.5 and 24.99 gm and over weighted if it exceeds 24.99 gm [41]. 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 a total 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) according to nutritional status which was evaluated using internationally accepted BMI guidelines[25].

The mean age in this study was 45.3(standard deviation (SD) 13.045) years overall and 47.77(SD 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 overweighted group.

Table-I: Characteristics of subjects in different glucose status groups

Variables	Malnourished (n=186)		Over-weighed (n=274)		Total(n=460)	
	Mean(SD)		Mean(SD)		Mea n(SD)	
Age(in years)	47.77(14.55)		43.62(11.64)		45.31(13.04)	
BMI(kg/m^3	16.56(1.38)		22.23(3.37)		19.94(4.11)	
Sex	No.	%	No.	%	No.	%
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=StandardSD=Standard deviation, BMI= Body mass index, History of Arsenic exposure.*

Those in the overweigthed group had a higher mean BMI than those in the malnourished groups in table 1. The chi-squared test indicated that there was a significant association between nutritional status and history of Arsenic exposure(P<0.001).

Moreover, table 1 shows that 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).

It is clear that one should specify the training and test dataset before conducting any training neural network architecture. Table 2 illustrates the nutritional status of the training and testing datasets of the sample.

Table-II : Distribution of nutritional status of the sample in the training and testing data sets.

Variable	Training datasets		Testing datasets		Total
	No.	%	No.	%	No.
Malnourished(BMI<18.5)	140	75	46	25	186
Overweighted(BMI>18.5)	206	75	68	25	274
Total	346	75	114	25	460

Comparative Study

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

Table-III: Odds ratios and coefficients of binary logistic regression analysis of factors associated with glucose status.

Characteristics	Coefficient	S.E.	OR	95.0% C.I.for EXP(B)	
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*

Table 4 shows the true and predicted status of subjects in the training and testing datasets as well as for all subjects. 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 (Table 5).

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

True Status	*Predicted Status using logistic -regression*		
	Malnourished No.	***Over-weighted No.***	***Total No.***
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

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

True Status	**Predicted Status using ANN architecture**		
	Malnourished No.	**Over-weighted No.**	**Total No.**
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

Discussion

Limited studies have indicated that poor nutritional status may increase the risk of arsenic related health effects[42-46]. Participants with poor nutritional status (weight below 80% of the standard body weight for their age and sex) were reported form West Bangal, 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[44]. 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[42,43,45,46]. 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.

Conclusion

In this study, we used the primary database of the patients 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. Thus we conclude that this study demonstrate a significant performance of artificial neural network than the binary logistic regression models in classification of malnourished (BMI < 18.5 gm) participants from over-weighted (BMI > 18.5 gm) ones.

References

1. Y. L. Eldon, "Artificial Neural Networks and Their Business Applications". Inf and Mang, 1994, 27(5): 303- 313.
2. S. S. Warren, "Neural Networks and Statistical Models", Preceedings of the 19the Annual SAS Users Group International conference, April 1994, USA.
3. Alonso-Betanzos A. 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. International journal of medical informatics, 2000, 57: 389-405.
6. Schwartz WB, Medicine and the computer: the promise and problems of change. New England journal of medicine, 1970, 283: 1257-64.
7. Shortliffe EH. The edolesence of Al in medicine: will the field come of age in the '90s? Artificial intelligence in medicine, 1993, 5(2):93-106.
8. Park J, Edington DE. A sequential neural network model for diabetes prediction. Artificial intelligence in medicine, 2001, 23(3): 277-93.
9. Ergun U. Classification of carotid artery stenosis of patients with diabetes by neural networks and logistic regression. Computers in biology and medicine, 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 in medicine, 1999, 16(30: 299-309.
13. Kennedy RL et. al. An artificial neural network system for diagnosis of AMI in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Computer methods and programs in biomedicine, 1997, 52(2): 93-103.
14. Cross SS et al. 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 et al. A Bayesian neural networks approach for modeling censored data with an application to prognosis after surgery for breast cancer. Artificial intelligence in medicine, 2003, 28(1): 1-25.
18. Tafeit E et al. The determination of three subcutaneous adipose tissue compartments in non-insulin-dependent diabetes mellitus women with artificial neural networks and factor analysis. Artificial intelligence in medicine, 1999, 17: 181-93.
19. Rahman, M., Tondel, M., Chowdhury, I. A. et al, Relations between exposure to arsenic, skin lesions, and glucosuria. Occupational Environmental Methodology, 199; 56, 277-281.
20. Bangla, P., Kaiser, J., India's spreading health crisis draws global arsenic experts. Science 1996; 274: 174-175.
21. DeFronzo, R. A., Bonadonna, R.C., Ferrannini, E., 1997. Pathogenesis of NIDDM. In: Alberti, K.G.M.M., Zimmet, P., DeFronzo, R.A., Keen, H. (Eds.), International Test book of Diabetes Mellitus, second ed. Wiley, New York, pp. 635-711.
22. Tseng, C. H., Tseng, C. P., Chiou, H. Y. et al. 2002. Epidemiologic evidence of diabetogenic effect of arsenic, Toxicology letters, 133, pp. 69-76.
23. Rahman M, Tondel M, ahmed SA, et al. Diabetes mellitus associated with arsenic exposure in Bangladesh. American Journal of Epidemiology, 1996;148: 196-203.

24. A. Kazemnejad, Z. Batvandi J. Faradmal. Comparison of artificial neural network and binary logistics regression for determination of impaired glucose tolerance/diabetes, Eastern Mediterranean Health Journal, vol. 16, No. 6. 2010.

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. "Generalised Obesity and Regional Adiposity in Adult White and Migrant Muslim Males from Pakistan in Peterborough." Journal of Royal Society for the Promotion of Health, 1996; 116: 161-167.

27. Deurenberg, P., Weststrate, A. J., Seidell, C. J. "Body mass index as a measure of body fatness: age- and sex-specific prediction formulas." British Journal of Nutrition 1991; 65: 105-114.

28. Ferro-Luzzi, A., Sette, S., Franklin, M., James, T. P. W. " A simplified approach of assessing adult chronic deficiency". European Journal of Clinical Nutrition 1992; 46: 173-186.

29. Garrow, J. S. and Webster, D. J. "Quetelets's index (W/H) as a measure of fatness." International Journal of Obesity, 1995; 9: 147-153.

30. James, T. P. W., Mascie-Taylor, N. G. C., Norgan et al. "The value of arm circumference measurements in assessing chronic energy deficiency in Third World adults." European Journal of Clinical Nutrition 1994; 48: 883-894.

31. Keys, A., Fidanza, F., Karvonen, M. J. et al. "Indices of relative weight and obesity." Journal of Chronic Disease 1972; 25: 329-43.

32. Khosla, T and Lowe, R. " Indices of obesity derived from body weight and height." British Journal of Preventive and Social Medicine 1967; 21: 122-128.

33. Lee, R. D. and Nieman, D. C. (2003). Nutritional Assessment. 2nd Edition, New York: McGraw Hill.

34. Lohman, T. O., Roche, A. F., Martorell, R. (1988). Anthropometric Standardization Reference Manual. Chicago: Human Kinetics Books.

35. Quetelet, L. A. J., (1842). " A treatise on man and the development of his faculties." In: Comparative statistics in the 19th century. Ediburgh: William and Robert Chambers.

36. Smalley, K. J., Knerr, A. N., Kendrick, Z. V. et al." Reassessment of body mass indices." American Journal of Clinical Nutrition, . 1990; 52,405-408.

37. Strain, G. W. and Zumoff, B. "The relation of weight-height indices of obesity to body fat content." Journal of American College of Nutritio 1991; 11,:715-718.

38. Das, S., and Bose, K. " Body mass index and chronic energy deficiency among adult Santals of Purulia District, West Bengal, India." International Journal of Human Sciences, 2010; 2(2): 488-503.

39. Abul H. M., Shahidullah, S. M., Smith, W.et al. Association between chronic Arsenic exposure and nutritional status among the women of child bearing age: A case-control study in Bangladesh." International journal of environmental research and public health, 2010;7: 2811-2821.

40. Abul H. M., Shahidullah, S. M., Smith, W.et al. Association between nutritional status and arsenicosis due to chronic arsenic exposure in Bangladesh." International journal of environmental research and public health, 2004;14: 99-108.

41. Shetty, P. S. Body Mass Index, a measurement of Chronic Energy Deficiency in Adults; FAO: Rome, Italy, 1994.

42. Borgono, J. M., Vicent, P., Venturino, H. et al " 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.

43. Hsueh, Y. M., Cheng, G. S., Wu, M. M., et al Multiple risk factors associated with arsenic-induced skin cancer: effects of chronic liver disease and malnutritional status, Brazilian Journal of Cancer, 1995, 71, 109-114.

44. Majumder, D. N., Haque, R., Ghosh, N., De, B. K et al.Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. International Journal of Epidemiology, 1998, 27: 871-877.

45. Tseng, W. P. Effects and dose-response relationships of skin cancer and blackfoot disease with arsenic. Environmental Health Perspective, 1977; 19: 109-119.

46. Zaldivar, R. Ecological investigations on arsenic dietary intake and endemic chronic poisoning in man: Dose-response curve. Zentralbl Bakteriol [orig B] 1977; 164: 481-484.