Comparison of artificial neural network and binary logistic regression for determination of impaired glucose tolerance/diabetes

A. Kazemnejad,[1] Z. Batvandi[1] and J. Faradmal[1]

مقارنة بين الشبكة العصبية الاصطناعية والتحوف اللوجستي الثنائي للتعرف على السكري وخلل تحمل الغلوكوز

انوشِروان كاظم نجاد، زيبا بتوندي، جواد فردمال

الخلاصة: قارن الباحثون بين نماذج مرتكزة على الشبكة العصبية الاصطناعية (المدرك المتعدد الطبقات) والتحوف اللوجستي الثنائي، من حيث قدرتهم على التمييز بين الأشخاص غير المصابين بالمرض وبين المصابين بالسكري أو بخلل تحمل الغلوكوز، الذين يشخصون بقياس الغلوكوز بعد الصيام، وقد جمع الباحثون المعطيات الديموغرافية (السكانية) والقياسات البشرية والسريرية من 7222 مشارك في الدراسة ممن تتراوح أعمارهم بين 30 و88 عاماً في دراسة السكر والشحوم في طهران. وبلغت القيمة الإحصائية لكابا كوهين 0.229 بالنسبة للتحوف اللوجستي، و0.218 بالنسبة للمدرك، وبلغت المساحة تحت المنحى ROC 0.760 بالنسبة للتحوف اللوجستي، و0.770 بالنسبة للمدرك. ولم يكن هناك فرق في الأداء بين النماذج استناداً إلى التحوف اللوجستي والشبكة العصبية الاصطناعية من حيث التفريق بين المصابين بخلل تحمل الغلوكوز والسكري وغير المصابين بالمرض.

ABSTRACT Models based on an artificial neural network (the multilayer perceptron) and binary logistic regression were compared in their ability to differentiate between disease-free subjects and those with impaired glucose tolerance or diabetes mellitus diagnosed by fasting plasma glucose. Demographic, anthropometric and clinical data were collected from 7222 participants aged 30–88 years in the Tehran Lipid and Glucose Study. The kappa statistics were 0.229 and 0.218 and the area under the ROC curves were 0.760 and 0.770 for the logistic regression and perceptron respectively. There was no performance difference between models based on logistic regression and an artificial neural network for differentiating impaired glucose tolerance/diabetes patients from disease-free patients.

Comparaison d'un réseau de neurones artificiels et de la régression logistique binaire dans la détermination de l'altération de la tolérance au glucose et du diabète

RÉSUMÉ Des modèles reposant sur un réseau de neurones artificiels (de type perceptron multicouche) et sur la régression logistique binaire ont été comparés. Ce parallèle portait sur leur capacité de différentiation entre sujets sains et individus présentant une altération de la tolérance au glucose ou un diabète sucré diagnostiqué par glycémie à jeun. Les données démographiques, anthropométriques et cliniques des 7 222 participants, âgés de 30 à 88 ans, de l'étude sur les lipides et le glucose réalisée à Téhéran ont été récupérées. Le test statistique Kappa de Cohen a permis d'obtenir des coefficients de 0,229 et 0,218 et les aires sous les courbes ROC étaient de 0,760 et 0,770 pour la régression logistique et le modèle de type perceptron, respectivement. Aucune différence n'a été constatée entre le modèle de régression logistique et celui reposant sur un réseau de neurones artificiels en termes de performance de distinction entre sujets sains et patients présentant une altération de la tolérance au glucose ou un diabète.

[1]Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Islamic Republic of Iran (Correspondence to A. Kazemnejad: kazem_an@modares.ac.ir).

Received: 25/02/08; accepted: 02/07/08

Introduction

Artificial intelligence has been proposed as a reasoning tool to support clinical decision-making since the earliest days of computing [*1–5*]. Artificial neural networks are a computer modelling technique based on the observed behaviours of biological neurons [*6*]. This is a non-parametric pattern recognition method which can recognize hidden patterns between independent and dependent variables [*7*].

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) [*8*]. 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 [*9*]. A typical multilayer perceptron is illustrated in Figure 1. Artificial neural networks have been applied to diagnosis and decision-making in various medical fields [*10–14*].

Statistical methods such as discriminant analysis and logistic regression have commonly been used to develop models for clinical diagnosis and treatment [*3*]. 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 [*15*], modelling for surgical decision-making for patients with traumatic brain injury [*3*] and survival of alcoholic patients with severe liver disease [*14*]. In contrast, others have reported that artificial neural networks and statistical models yielded similar results [*7,16*].

Diabetes mellitus (DM) is a common chronic disease in the adult population and is associated with a significantly increased risk of micro- and macrovascular disease. DM is frequently insidious in onset and patients may be relatively symptom-free for years before diagnosis. In the Islamic Republic of Iran, there are about 3 million individuals affected by DM and with increasing urbanization, the prevalence of DM is rising rapidly. There is thus an urgent need to identify and manage patients with DM, especially in groups at higher risk for the disease and its complications [*17*].

In this study, we developed a multilayer perceptron artificial neural network to differentiate between disease-free subjects and those with impaired glucose tolerance (IGT) or DM and compared the accuracy of this model with the more traditional method of binary logistic regression for the prediction of patients' glucose metabolism status.

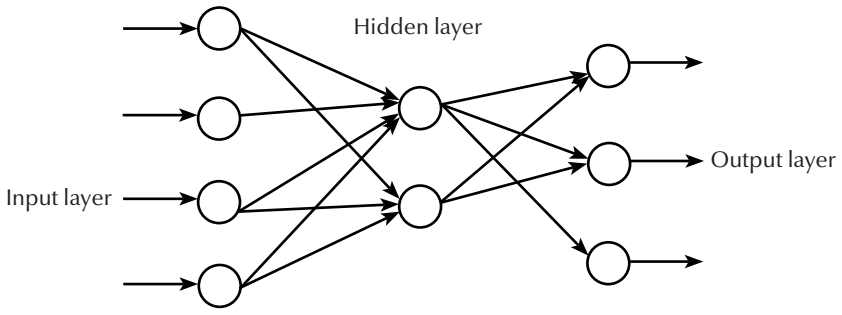

Figure 1 Typical multilayer perceptron model with 4 neurons in the input layer, 2 in the hidden layer and 3 in the output layer and with no direct connection from input to output layers

Methods

Study population

The data for the study were obtained from the database of the Tehran Lipid and Glucose Study (TLGS), which was conducted to determine the risk factors for atherosclerosis among Tehran's urban population, to develop population-based measures to change the lifestyle of the population and to prevent the rising trend of DM and dyslipidaemia [*18*]. For the TLGS, cluster random sampling was used to recruit 15 000 people from the 13th district of urban Tehran, the capital of the Islamic Republic of Iran. Among this population, 7222 adults aged 30–88 years (43.2% male and 56.8% female) who had no prior record of DM and had complete information were the subjects of the present study. Data were collected at the TLGS clinic between February 1999 and August 2001.

Patients' demographic and clinical characteristics

Fasting plasma glucose (FPG) level was used to classify the glucose metabolism status of each subject according to American Diabetes Association (ADA) criteria [*19*]. A blood sample was drawn into vacutainer tubes between 07:00 and 09:00 hours from all study participants after a 12–14 hour overnight fast. Subjects were classified as: normal glucose or disease-free (FPG < 110 mg/dL); IGT (FPG ≥ 110 < 126 mg/dL); or diabetic (FPG ≥ 126 mg/dL).

The demographic and clinical data used as predictors in the models were: patient's age, body mass index (BMI), waist-to-hip ratio (WHR), history of hypertension and history of diagnosis of hyperlipidaemia. Hypertension was defined as any prior diagnosis of hypertension by a physician or if the patient was taking antihypertensive medication at the time of interview or in the previous 1 month. Weight and height were measured according to standard protocols. BMI was calculated by dividing the weight (kilogram) by the

square of height (metres). WHR was the waist circumference measured at the level of the umbilicus divided by the hip circumference measured over light clothing at the widest girth of the hip.

Prediction models

We applied 2 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 3-layer topology (input, hidden and output layers) with 4 neurons in the hidden layer and no direct connection from the input to output layers [9]. Given enough hidden nodes and sufficient data, it can approximate any function to any desired degree of accuracy. The error back-propagation learning algorithm is a powerful approach and, despite its slow convergence, is one of the most popular and successful algorithms for pattern recognition.

The 2 different models were compared in their ability to predict glucose metabolism status from the patients' demographic and clinical data. To do this we first merged the subjects in the DM and IGT groups. Then we split the database into 2 groups: a training dataset containing approximately 75% of the sample and a testing dataset containing approximately 25% of subjects. The training dataset was used to develop the logistic regression and perceptron models by introducing the disease status of subjects (according to ADA criteria) into the models. The testing dataset was used by the models for predicting the glucose tolerance status of subjects.

Comparison tools were the kappa measurement of agreement and the area under the receiver operating characteristics (ROC) curve. The ROC curve was obtained by plotting 1 minus the specificity rate against the sensitivity rate for all possible cut-off points.

Software

The neural network development software used in this study was *R*, version 2.5.0 package (nnet version 7.2-290) (*R* is an open-source system available at http://www.r-project.org). Other statistical analyses, including descriptive statistics and analysis of variance (ANOVA) to compare mean values and the binary logistic regression, were performed using *SPSS*, version 13.0.

Results

Patients' clinical characteristics

Among 7222 participants aged 30 years or over, 629 (8.7%) suffered from DM, 418 (5.8%) had IGT and the remainder were disease-free by ADA criteria.

The mean age in this study was 47.7 [standard deviation (SD) 12.5] years overall and 46.4 (SD 12.3) years for the disease-free group (Table 1). One-way ANOVA indicated that the mean age of the 3 groups was significantly different ($P < 0.001$) and Tukey *post hoc* multiple comparison test showed that the disease-free group was younger than the DM ($P < 0.001$) and IGT patients ($P < 0.001$).

Those in the disease-free group had a lower mean BMI than those in the DM ($P < 0.001$) and IGT groups ($P < 0.001$) Table 1. The lowest and the highest WHR were 0.56 and 1.45 respectively. Subjects in the DM and IGT groups had higher WHR

Table 1 Characteristics of subjects in different fasting plasma glucose status groups

Variable	Disease-free (*n* = 6175)		IGT (*n* = 418)		DM (*n* = 629)		Total (*n* = 7222)
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)
Age (years)	46.4 (12.3)		52.9 (12.0)		57.0 (10.2)		47.7 (12.5)
Anthropometric measures							
BMI (kg/m²)	27.3 (4.5)		29.5 (4.9)		28.8 (4.6)		27.6 (4.6)
WHR	0.88 (0.08)		0.93 (0.08)		0.94 (0.08)		0.89 (0.09)
	No.	%	No.	%	No.	%	No.
Sex							
Male	2704	43.8	179	42.8	239	38.0	3122
Female	3471	56.2	239	57.2	390	62.0	4100
History of hyperlipidaemia							
Yes	1379	22.3	152	36.4	321	51.0	1852
No	4796	77.7	266	63.6	308	49.0	5370
History of hypertension							
Yes	931	15.1	128	30.6	239	38.0	1298
No	5244	84.9	290	69.4	390	62.0	5924

SD = standard deviation.
IGT = impaired glucose tolerance; DM = diabetes mellitus; BMI = body mass index; WHR = waist-to-hip ratio.

than those in the disease-free group (Table 1).

The chi-squared test indicated that there was a significant association between glucose tolerance status and history of hyperlipidaemia ($P < 0.001$). Table 1 shows that the IGT and particularly the DM groups had a higher proportion of subjects with a positive history of hyperlipidaemia compared with the disease-free group (36.4%, 51.0% and 22.3% for the IGT, DM and disease-free groups respectively). Participants with DM or IGT were more likely to have a positive history of hypertension than those diagnosed as disease-free (30.6%, 38.0% and 15.1% for IGT, DM and disease-free groups respectively). The association between glucose tolerance status and history of hypertension was significant ($P < 0.001$).

Table 2 illustrates the glucose tolerance status of the training and testing datasets of the sample.

Table 2 Distribution of fasting plasma glucose status of the samples in the training and testing datasets

Variable	Training dataset		Testing dataset		Total
	No.	%	No.	%	No.
Disease-free	4673	75.7	1502	24.3	6175
IGT or DM	802	76.6	245	23.4	1047
Total	5475	75.8	1747	24.2	7222

IGT = impaired glucose tolerance; DM = diabetes mellitus.

Comparison of models

Using binary logistic regression all factors were significantly associated with glucose tolerance status (Table 3). Age, sex, BMI and WHR were significant risk factors for DM. Meanwhile, those who were suffering from hyperlipidaemia or hypertension had a higher risk of DM and IGT.

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 72.2% of cases with IGT or DM in the training dataset, 71.0% in the testing set and 71.9% of all subjects. The area under the ROC curve for this model was 0.760 and the kappa statistic was 0.229, showing that the emerged classification was not due to chance ($P < 0.001$).

The sensitivities of the perceptron for the training and testing datasets and for all subjects were 79.4%, 77.1% and 78.9% respectively (Table 5). These values were obtained using 0.136 as the cut-off point. Based on Table 5, the specificities of ANN for the training, testing and total of the dataset were 62.2%, 59.4% and 61.5% respectively. kappa statistic was 0.218 which was significantly different from zero. The area under the ROC curve for this model was 0.770.

Discussion

In this study, we used the TLGS database to develop models to try to distinguish patients with IGT or DM from disease-free patients. The accuracy of the perceptron and binary logistic egression models in predicting a subject's glucose tolerance status were compared using the kappa statistic and the area under the ROC curve. The kappa value for logistic regression (0.229) was slightly higher than for the perceptron (0.218). Although the kappa values were significantly different from zero, they were far from 1. The small number of covariates may be responsible for the low kappa values and the large number of subjects may be the cause of the significance. Therefore in terms of the kappa statistic the neural network model did not perform better than binary logistic regression. Also, the area under the ROC curve was barely different in the 2 models (0.760 for logistic regression and 0.770 for perceptron). The 2 models not only resulted in almost the same confusion matrix for the training dataset, but also for the testing dataset.

For binary logistic regression, a good model depends on determining the relation of the mean response (or logit function of it) to the predictor(s). But it is sometimes difficult to guess the appropriate form for this relationship. Nevertheless, logistic regression can identify the effect and the direction of each factor on the (mean) response.

On the other hand, artificial neural networks are useful tools for prediction

Table 3 Odds ratios and coefficients of binary logistic regression analysis of factors associated with glucose tolerance status

Characteristic	Coefficient	SE	OR	95% CI
Intercept	–9.954	0.531**	–	–
Sex (male)	–0.244	0.098*	0.78	0.65–0.95
Age	0.040	0.004**	1.04	1.03–1.05
History of hyperlipidaemia	0.573	0.085**	1.77	1.50–2.10
History of hypertension	0.247	0.098*	1.28	1.06–1.55
BMI	0.061	0.010**	1.06	1.04–1.08
WHR	4.680	0.611**	1.60	1.42–1.80[a]

*P < 0.05; **P < 0.01.
[a]Computed for 0.1 increase in WHR.
SE = standard error; OR = odds ratio; CI = confidence interval; BMI = body mass index; WHR= waist-to-hip ratio.

Table 4 Number of correct diagnoses of glucose tolerance status using binary logistic regression model

True status	Predicted status using logistic regression		
	Disease-free	IGT or DM	Total
	No.	No.	No.
Training dataset			
Disease-free	3148	1525	4673
IGT or DM	223	579	802
Total	3371	2104	5475
Testing dataset			
Disease-free	985	517	1502
IGT or DM	71	174	245
Total	1056	691	1747
Overall			
Disease-free	4133	2042	6175
IGT or DM	294	753	1047
Total	4427	2795	7222

IGT = impaired glucose tolerance; DM = diabetes mellitus.

Table 5 Number of correct diagnoses of glucose tolerance status using multilayer perceptron model

True status	Predicted status using perceptron		
	Disease-free	IGT or diabetic	Total
	No.	No.	No.
Training dataset			
Disease-free	2907	1766	4673
IGT or DM	165	637	802
Total	3072	2403	5475
Testing dataset			
Disease-free	892	610	1502
IGT or DM	56	189	245
Total	948	799	1747
Overall			
Disease-free	3799	2376	6175
IGT or DM	221	826	1047
Total	4020	3202	7222

IGT = impaired glucose tolerance; DM = diabetes mellitus.

when the form of the relation is unknown. Determining the factor contributions in artificial neural networks models, however, is intrinsically difficult. Unlike traditional statistical models, neural networks do not help in identifying the most statistically influential input factor. The complexity of neural networks makes it difficult to relate their output to input. Hart and Wyatt argued that this "black box" aspect is the major barrier to the acceptance of neural networks for medical decision systems [*20*]. If prediction is the only objective, then neural network models provide acceptable results whereas binary logistic regression could also identify the effect of factors on the classification.

We conclude that this study did not demonstrate a significant performance difference between models based on logistic regression and an artificial neural network for differentiating IGT and DM patients from disease-free ones.

References

1. 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.
2. Lisboa PJA. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural networks*, 2002, 15:11–39.
3. Li YC, Chiu WT, Jian WS. Neural networks modeling for surgical decisions on traumatic brain injury patients. *International journal of medical informatics*, 2000, 57:389–405.
4. Schwartz WB. Medicine and the computer: the promise and problems of change. *New England journal of medicine*, 1970, 283:1257–64.
5. Shortliffe EH. The adolescence of AI in medicine: will the field come of age in the '90s? *Artificial intelligence in medicine*, 1993, 5(2):93–106.
6. Park J, Edington DE. A sequential neural network model for diabetes prediction. *Artificial intelligence in medicine*, 2001, 23(3):277–93.
7. Ergun U et al. Classification of carotid artery stenosis of patients with diabetes by neural networks and logistic regression. *Computers in biology and medicine*, 2004, 34:389–405.
8. Rosenblatt F. *The perceptron: a perceiving and recognizing automation. Cornell Aeronautical Laboratory report 85-460-1*. Ithaca, New York, Cornell Aeronautical Laboratory, 1957.
9. Bishop CM. *Neural networks for pattern recognition*. Oxford, Oxford University Press, 1995.
10. 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(3):299–309.
11. 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.
12. 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.
13. Dybowski R, Gant V. Artificial neural network in pathology and medical laboratories. *Lancet*, 1995, 346:1203–7.
14. Lapuerta P, Rajan S, Bonacini M. Neural networks of outcomes in alcoholic patients with severe liver disease. *Hepathology*, 1997, 25:302–6.
15. 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.
16. 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.
17. Azizi F et al. Distribution of blood pressure and prevalence of hypertension in Tehran adult population: Tehran Lipid and Glucose Study (TLGS), 1999–2000. *Journal of human hypertension*, 2002, 16(5):305–12.
18. Azizi F, Rahmani M, Emami H. Tehran Lipid and Glucose Study: rationale and design. *CVD prevention*, 2000, 3:242–7.
19. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes care*, 1997, 20:1183–97.
20. Hart A, Wyatt J. Evaluating black boxes as medical decision-aids: issues arising from a study of neural networks. *Medical informatics*, 1990, 15:229–36.

Note from the Editor

We wish to draw the kind attention of our potential authors to the importance of applying the editorial requirements of EMHJ when preparing their manuscripts for submission for publication. These provisions can be seen in the Guidelines for Authors, which are available online at http://www.emro.who.int/emhj.htm, and are published at the end of the first issue of each volume. We regret that we are unable to consider papers that do not conform to the Guidelines.