TÜBİTAK

ORIGINAL ARTICLE

Turk J Med Sci
2009; 39 (1): 5-12
© TÜBİTAK
E-mail: medsci@tubitak.gov.tr
doi:10.3906/sag-0709-31

Artificial Neural Network Analysis for Prediction of Headache Prognosis in Elderly Patients

Bahar TAŞDELEN[1]
Sema HELVACI[1]
Hakan KALEAĞASI[2]
Aynur ÖZGE[2]

[1] Department of Biostatistics,
Faculty of Medicine,
Mersin University,
Mersin - TURKEY

[2] Department of Neurology,
Faculty of Medicine,
Mersin University,
Mersin - TURKEY

Aim: To investigate the ability of neural networks to detect and classify the complete improvement of headache in elderly patients during the follow- up period.

Materials and Methods: The multilayer perceptron (MLP), which is the most common neural network, was used to predict prognosis of headache in elderly patients. The data set was divided into training and test sets, and back-propagation algorithm was used as the learning algorithm. The accuracies of the models to predict completely improved patients at the end of 20, 40, and 60 months of follow-up were evaluated by means of the areas under the receiver operating characteristic (ROC) curves.

Results: The classification results showed the neural network models had good performance in both training and test phases. In addition, the areas under the ROC curve for each period showed that the accuracies of the models to predict the completely improved patients were in the interval of 0.75-0.90. Conclusions: Neural network model for grouped survival data can be used as a prognostic model. If the prevalence of a disease is low, the sensitivity of the model for detection of the patients with disease will be low.

Key Words: Artificial neural networks, headache, multilayer perceptron, prognosis

Yaşlı Hastalarda Baş Ağrısı Prognozunun Tahmini İçin Yapay Sinir Ağları Analizi

Amaç: Çalışmanın amacı, sinir ağlarının takip süresi içinde başağrısı tam olarak iyileşen yaşlı hastaları belirleyerek sınıflandırma performansını incelemektir.

Gereç ve Yöntem: Yaşlı hastalarda başağrısı prognozunu tahmin etmek için, en yaygın sinir ağı olan çok tabakalı perseptron kullanıldı. Veri seti, eğitim ve test setlerine ayrılarak, öğrenme algoritması olarak geriye yayılım algoritması kullanıldı. Tedavinin 20, 40 ve 60. aylarında hastalardaki tam iyileşmeyi tahmin etmek için kullanılan modellerin doğruluk dereceleri ROC eğrileri altında kalan alanlar kullanılarak değerlendirildi.

Bulgular: Sınıflandırma sonuçları, sinir ağı modellerinin hem eğitim hem de test aşamalarında yüksek performansa dahip olduklarını göstermiştir. Ayrıca, her tedavi periyodu için elde edilen ROC eğrisi altında kalan alanlar, tam iyileşmeyi tahmin etmek için kullanılan modellerin doğruluk derecelerinin 0.75-0.90 aralığında olduğunu göstermiştir.

Sonuç: Gruplandırılmış sağkalım verileri için sinir ağı modeli prognostik model olarak kullanılabilir. Eğer bir hastalığın görülme sıklığı düşük ise, modelin gerçek hastaları tayin etme gücü de (duyarlılığı) düşük olacaktır.

Anahtar Sözcükler: Yapay sinir ağları, baş ağrısı, çok tabakalı perseptron, prognoz

Received: September 24, 2007
Accepted: August 27, 2008

Correspondence
Bahar TAŞDELEN
Department of Biostatistics,
Faculty of Medicine,
Mersin University,
Mersin - TURKEY

bahartasdelen@mersin.edu.tr

Introduction

Prognostic models are used for diagnosis and analysis of survival time in many clinical studies. A number of statistical methods have been developed for these studies. Logistic regression (LR) is one of them and is used to estimate the probability of an outcome. In the model, covariate may be a prognostic score, or the model consists of a time-dependent outcome variable. In such cases, survival analysis is used to determine the change in the survival ratio across time. Censored data can be thought of as an important task in survival analysis. The data that has not been observed during the data collection period is referred to as “censored” data, and Cox regression is the method used for its evaluation (1).

Recently, artificial neural networks (ANNs) have become popular as diagnostic and prognostic models (2-12). They have been applied to diagnose disease and predict the survival ratio of patients. However, for medical analysis, ANNs have been shown to have some disadvantages as well as advantages. The most important advantages of ANN are their discriminating power, detection of complex and nonlinear relationships between independent and dependent variables, and prediction of the case. On the other hand, ANNs can be considered as "black box" methods, the model is developed empirically, they can be over-fitted for training data, and their usage is very difficult because of computational requirements (3,13,14). The performance of an ANN depends on the number of parameters, the network weights, the selection of an appropriate training algorithm, the type of transfer functions used, and the determination of network size. Another disadvantage of using ANNs is that they require the initialization and adjustment of many individual parameters to optimize their classification performance (2). However, some software for ANNs such as STATISTICA Neural Networks Module includes Intelligent Problem Solver, which does not demand initialization and finds optimum network for the data.

Many researchers have compared ANN versus LR, discriminant analysis and Cox proportional hazards model. Some of them found that ANN and LR have similar classification performance (15,16), while others showed that the differences in performances of ANN and Cox proportional hazards model for predictions were not significant (4-7). In LR, the most parsimonious model is the best model. If the model involves a large number of variables relative to the number of subjects, unrealistically large coefficients and/or standard errors are estimated (17). Hence, the model complexity is reduced by performing variable selection. Compared to LR, neural network models are more flexible. ANN can be seen as the nonlinear generalization of LR (13).

Headache is an important cause of morbidity and loss of productivity and is one of the most common complaints in general practice for all age groups (18-20). After the age of 65 years, more than 13% of women and 7% of men continue to complain of headache, and this causes various social and healthcare problems (3). There are few data about the determination of headache prognosis in these age groups. Primary headaches are the most frequently seen headache form in the elderly, and the knowledge about the clinical types and the natural history of these headaches is quite limited. Hence, analysis of the time for the total improvement of patients, especially those with primary headaches, is very important. Moreover, periodic investigation of this time and determining which prognostic factor is more efficient in which period will be meaningful to select a suitable treatment. ANN can be used to solve this problem by grouping improvement time.

The aim of this study was to investigate the usefulness of neural networks to predict and classify the complete improvement of headache in elderly patients.

Materials and Methods

Patients

The study population included 341 patients who had admitted to the headache clinic of Mersin University Hospital between March 1999 and January 2005. Headache diagnosis was reviewed by the same headache specialist (AO) according to the International Classification of Headache Diagnosis second edition (ICHD-2) criteria (21) and subgroup distributions are reported in Table 1. While following these patients, the results of prophylaxis treatments and attack treatments were observed periodically. The follow-up periods were 1-3 months and treatments were changed or some consultations were repeated in those patients who did not show meaningful response to treatment in one period. For example, the psychiatry consultations were repeated for the patients with migraine if no significant variations of their pain characteristics were determined in six-month prophylaxis treatment. A simultaneous treatment was applied to patients who had comorbid disorders. Finally, invasive procedures were used for patients who had no response to any treatment, if they accepted. The most important factors for the low treatment success were adaptation difficulty of disease, disregard of headache drugs for the others taken for comorbid diseases, and terminating treatment without informing the physician.

The headache features were recorded using headache diaries and included the duration, frequency and severity during follow-up. The patients were seen monthly during the first three months, and once every two months in the following months. In each visit, the patients and the headache diaries were evaluated by a headache specialist, and the headache diagnosis and treatment options were revised, if needed. Headache severity was evaluated using visual analogue scale (VAS) (22). The response to treatment was evaluated in four categories as complete

Table 1. The number of patients in the subgroups of headache type.

Primary headache disorders	n	Secondary headache disorders	n
Migraine with aura	16	Acute post-traumatic headache	4
Migraine with visual aura	1	Chronic post-traumatic headache	17
Infrequent episodic tension-type headache	23	Headache attributed to TIA or ischemic stroke	9
Frequent episodic tension-type headache	71	Temporal arteritis	1
Chronic tension type headache	147	Carotidynia	1
Probable tension type headache	1		
aroxysmal hemicrania	3	Headache attributed to intracranial tumors	5
Idiopathic stabbing headache	3	Medication-overuse headache	1
Cough headache	5		
		Headache attributed to systemic infections	1
		Headache attributed to arterial hypertension	13
		Headache attributed to other metabolic disorders	1
		Headache attributed to disorders of neck	12
		Headache attributed to disorders of eye	3
		Headache attributed to major depressive disorders	1
		Headache attributed to generalized anxiety disorders	1
		Trigeminal neuralgia	4

TIA: transient ischemic attack

improvement, partial improvement, unchanged, and worsened. The frequency, severity and duration of headache were recorded in a special database at each visit. Complete improvement represents having no headache or a decrease in headache severity and duration of more than 80%. Partial improvement represents 30-80% decrease in headache severity and duration. Worsened represents increased headache severity or duration of more than 10% during the follow-up process.

In this study, inputs were sex (1: female, 2: male), age (years), follow-up time (months), headache type (1: primary, 2: secondary), headache duration (hours/day), headache frequency (days/month), and headache severity score (1-10). Binary output variable was recorded according to complete improvement before the end of the follow-up period. The patients improved in a given time were coded as 1, and the others as 0. Although we had four equal time intervals between 0 and 80 months, the distribution of patients to the response variable in the last period was the same as in the third period. Therefore, we used three neural network models for three response variables (complete improvement: less than 20 months, less than 40 months, and less than 60 months). The characteristics of input variables and the numbers of the patients who improved completely according to comorbidities are given in Tables 2 and 3.

Artificial Neural Networks

An ANN is a paradigm that uses interconnected artificial neurons and mathematical models in order to represent complex and nonlinear relationships between input and output variables. ANN architecture divides data into training (257 samples; 75%) and test (84 samples; 25%) parts.

Although different neural network types have been developed, the most common neural network is the multilayer perceptron (MLP), which is a feed-forward neural network type and generally uses back-propagation algorithm to develop a model to illustrate relationships between inputs and a desired output for training data. This model is then used to produce output for test data. The graphical illustration of the network is shown in Figure 1.

Table 2. The characteristics of the input variables.

		End of follow-up time (months)							
	Total	20		40		60		80	
Input variable		1[†]	0	1	0	1	0	1	0
Sex									
Female	258	19	239	22	236	25	233	25	233
Male	83	3	80	4	79	5	78	5	78
Age									
65-74	283	16	267	20	263	23	260	23	260
75-84	51	6	45	6	45	7	44	7	44
85-94	7	0	7	0	7	0	7	0	7
Headache type									
Primary	267	19	248	22	245	25	242	25	242
Secondary	74	3	71	4	70	5	69	5	69
Headache frequency									
Mean	20.09	17.09	20.29	17.92	20.27	16.70	20.41	16.70	20.41
Standard deviation	11.00	11.44	10.96	11.44	10.96	11.24	10.94	11.24	10.94
Headache duration									
Mean	13.96	11.68	14.12	11.62	14.15	11.40	14.21	11.40	14.21
Standard deviation	16.42	21.19	16.07	19.66	16.14	18.52	16.21	18.52	16.21
Headache severity									
Mean	6.18	5.91	6.20	5.92	6.20	5.87	6.21	5.87	6.21
Standard deviation	1.70	1.79	1.69	1.79	1.69	1.76	1.69	1.76	1.69

Table 3. The observed frequencies of complete improvement according to comorbidities.

		End of follow-up time (months)					
	Total	20		40		60	
		1[†]	0	1	0	1	0
Hypertension							
+	166	14	152	17	149	19	147
-	175	8	167	9	166	11	164
Diabetes Mellitus							
+	50	2	48	4	46	4	46
-	291	20	271	22	269	26	265
Coronary Artery Disease							
+	49	2	47	3	46	3	46
-	292	20	272	23	269	27	265
Lipidemia							
+	37	2	35	3	34	3	34
-	304	20	284	23	281	27	277
Stroke							
+	12	1	11	1	11	1	11
-	329	21	308	25	304	29	300
Nausea							
+	133	9	124	11	122	13	120
-	208	15	193	17	191	19	189
Vomiting							
+	27	1	26	1	26	1	26
-	314	21	293	25	289	29	285
Photophobia							
+	60	6	54	7	53	8	52
-	281	16	265	19	262	22	259
Phonophobia							
+	113	10	103	11	102	12	101
-	227	12	215	15	212	18	209

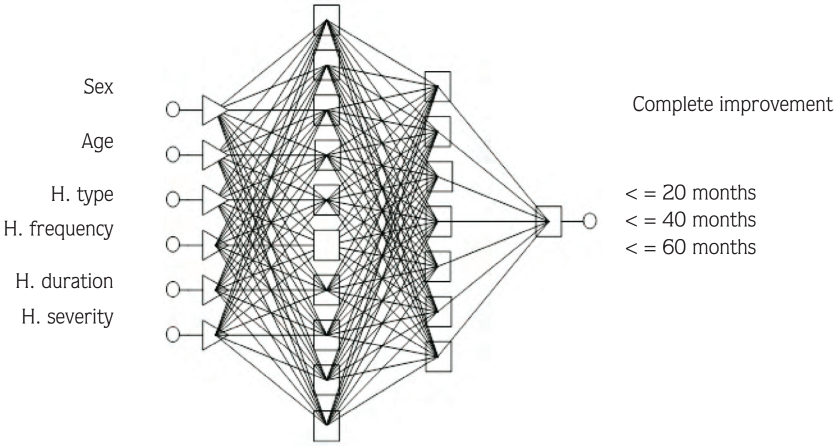

Figure 1. The neural network for prediction of complete improvement of patients with headache in given follow-up periods.

Training of the network requires minimization of the learning function, which is a distance measure between observed and predicted outputs. This procedure is applied iteratively. A MLP consists of three layers: an input layer, one or more hidden layers and an output layer. A MLP model that consists of a single hidden layer is

$$y_k = f\left(\alpha_k + \sum_{h=1}^{H} \omega_{hk} f\left(\alpha_h + \sum_{i=1}^{I} w_{ih} x_i\right)\right)$$

where I is the number of inputs, H is the number of hidden layers, α_H and α_k are the bias terms for hidden and output layers. The transfer function f is the logistic function for MLP

$$f(x) = \frac{\exp(x)}{1 + \exp(x)}$$

and the cross-entropy error function is used to classify binary output (5).

Results

Neural Networks Module of STATISTICA 7.0 software was used to develop the ANN model. A summary of the model is reported in Table 4. The model profile gives the model's structure, which is the form I:N-N-N:O, where I is the number of inputs, O is the number of outputs, the first N is the number of units in input layer, the second N is the number of units in hidden layer, and the last N is the number of units in output layer. The sensitivities of input variables for each model were given with their ratio and ranking values where the ratio for an input variable is the sensitivity of modeling performance that occurs if that variable is no longer available to the model. Since the relationships between comorbid disorders and complete improvement were not statistically significant (P > 0.05), ANN models did not include any of them. The classification results of the training and test phases of each model are given in Table 5.

In this study, we assumed that true positive was the prediction of a patient who improved completely as improved. Since the proportion of patients who improved completely was less than 0.10, the sensitivities (0.33, 0.39, and 0.50 for training samples; 0.25, 0.00, and 0.00 for test samples; 0.47, 0.37, and 0.36 for overall data) of all models were less than their specificities (0.99, 1.00, and 0.99 for training samples; 0.99, 0.97, and 0.97 for test samples; 0.99, 0.99, and 0.99 for overall data). Positive predictive values for each period in training samples and overall data set were detected between 0.67 and 1.00. Negative predictive values for each period in training samples and overall data set were found between 0.93 and 0.96.

Discussion

We trained MLP models and selected one of them with the best performance value for each output variable. The model used to predict complete improvement before 20 months includes six variables at input layer, two hidden layers with 10 and 8 units in each layer, and one output layer. In practice, we may predict which patient will improve completely, according to the characteristics of

Table 4. A detailed summary of multilayer perceptron models.

	End of follow-up time (months)		
	20	40	60
Profile	6:6-10-8-1:1	6:6-10-10-1:1	6:6-10-8-1:1
Train performance	0.945	0.957	0.929
Test performance	0.952	0.881	0.905
Sensitivity: Ratio (ranking)			
Sex	1.315 (3)	1.069 (2)	1.370 (2)
Age	1.095 (6)	1.030 (4)	1.225 (4)
Headache type	1.137 (5)	1.051 (3)	1.150 (5)
Headache frequency	1.639 (1)	1.348 (1)	2.197 (1)
Headache duration	1.149 (4)	0.912 (6)	1.305 (3)
Headache severity	1.454 (2)	1.004 (5)	1.057 (6)

Table 5. Confusion matrixes for training, test and overall sets. 1: Complete improvement, 0: Not complete improvement. The columns show observed frequencies and the rows show predicted frequencies for each period.

		Sampling					
		Training		Test		Overall	
End of follow-up time (months)		0	1	0	1	0	1
	0	237	12	79	3	316	15
20	1	2	6	1	1	3	7
	0	239	11	74	8	313	19
40	1	0	7	2	0	2	7
	0	231	16	76	6	307	22
60	1	2	8	2	0	4	8

sex, age, headache type, headache duration, headache frequency, and headache severity. In addition, we may want to learn which patient will improve completely at the end of a given follow-up period. Thus, we can use ANN to classify patients as improved or not completely improved at the end of a period by input variables.

Many researchers have compared ANN versus LR, discriminant analysis and classification and regression trees, and showed that ANN and LR have similar predictive performance [15]. On the other hand, the sensitivity and specificity of ANN modeling was superior to multivariate LR analysis [16]. The results of the comparison of the performance of multiple discriminant analysis (MDA), LR and ANNs showed that LR and MDA were both more efficient in the use of computer time than ANN. The results also suggested that the classification performance of all methods were almost the same (23). The accuracies of ANN model and Cox proportional hazards model with respect to their sensitivities, specificities, positive and negative predictive values and the areas under the receiver operating characteristic (ROC) curves were calculated; both models had similar performance (4). ANN models have been recently developed for the analysis of censored survival data (5-7).

This study gives important results about classification of patients in a given follow-up period. The performances of the model were high in both training and test phases for all periods. We noticed that the sensitivity of headache frequency was higher than the others for each follow-up period. In addition, sex was the common variable to predict improvement for each period. While all variables except headache duration in the second period were important in predicting, headache severity in the first period, headache type in the second period, and headache duration in the third period were the most important variables. Therefore, total improvement of headache before 20 months more likely depends on headache frequency, severity, and duration as well as on gender (Table 4).

The classification results showed that the neural network models for each period were good at discriminating patients who had not improved completely at the end of the given period. The areas under the ROC curve for each period are given in Figure 2. The accuracies of the models to predict completely improved patients at the end of 20, 40, and 60 months of follow-up were 0.895, 0.749, and 0.825, respectively.

Artificial neural networks can be used to analyze the complex and huge data sets since they are distribution-free models. They can be applied to multivariate nonlinear problems. In addition, they can detect complex nonlinear relationships between independent and dependent variables and interactions among estimating variables.

In medical practice, ANNs are generally used to diagnose and monitor the prognosis of a disease. Medical usage areas of ANN can be explained with some examples from the literature. ANNs have been used to determine prognosis in trauma, prognosis after cardiopulmonary resuscitation, outcome of treatment for ovarian cancer, prognosis in acquired immunodeficiency syndrome (AIDS), predicting mortality of patients with end-stage liver disease, prognosis for patients with colonic cancer, detecting extensive coronary artery disease, predicting length of stay in the intensive care unit following cardiac surgery, and predicting the risk of death for small-cell lung cancer patients (2,4,8,11,15,24-27). In this study, we also used ANN to predict complete improvement of elderly patients with headache.

Headache is an important cause of morbidity and loss of productivity and is one of the most common complaints in general practice. We thus used ANNs to establish prognosis of complete improvement of headache in patients over 65 years of age. Although ANN models have the best performance in large data sets, we notice that they might be good at predicting complete improvement of patients in a given follow-up period.

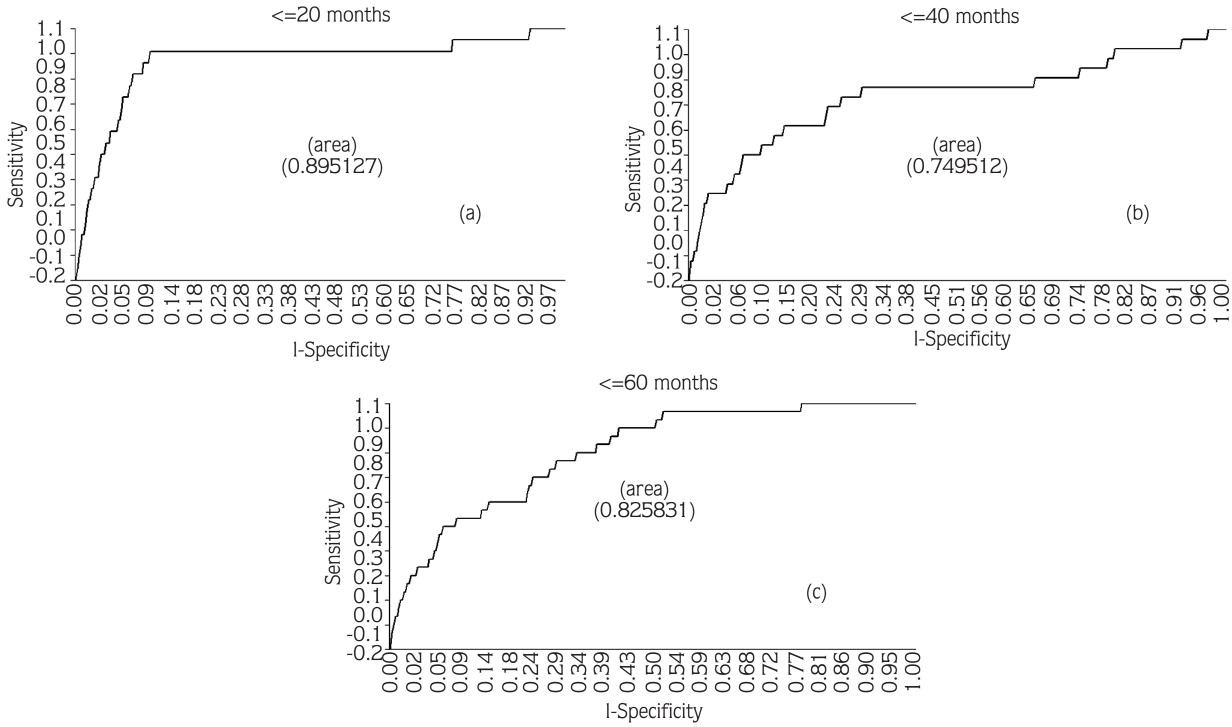

Figure 2. ROC curves and their accuracies for prediction of complete improvement in three periods after treatment.

Therefore, the neural network model for grouped survival data can be used as a prognostic model, and the significant risk factors can be determined using sensitivity analysis. In this study, the prevalence of complete improvement of patients over 65 years is low, and the sensitivity of the model for detection of patients who will improve completely at the end of the follow-up period is also low. Nevertheless, we can use ANN models to learn which risk factors affect the complete improvement of elderly patients with headache in a given period.

References

1. Abu-Hanna A, Lucas PJF. Prognostic models in medicine: artificial intelligence and statistical approaches. Method Inform Med 2001; 40: 1–5.
2. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. Mol Cancer 2005; 4: 29.
3. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. Stat Med 2000; 19: 541-61.
4. Ohno-Machado L. A comparison of Cox proportional hazards and artificial neural network models for medical prognosis. Comput Biol Med 1997; 27(1): 55-65.
5. Biganzoli E, Borrachi P, Mariani L, Marubini E. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Stat Med 1998; 17: 1169-86.
6. Biganzoli E, Borrachi P, Marubini E. A general framework for neural network models on censored survival data. Neural Netw 2002; 15: 209-18.
7. Eleuteri A, Tagliaferri R, Milano L, De Placido S, De Laurentis M. A novel neural network-based survival analysis model. Neural Netw 2003; 16: 855-64.
8. Allison JS, Heo J, Iskandrian AE. Artificial neural network modeling of stress single-photon emission computed tomographic imaging for detecting extensive coronary artery disease. Am J Cardiol 2005; 5: 178-81.
9. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. A comparison of statistical learning methods on the GUSTO database. Stat Med 1998; 17(21): 2501-8.
10. Keijsers NLW, Horstink MWIM, van Hilten JJ, Hoff JI, Gielen CCAM. Detection and assessment of the severity of levodopa-induced dyskinesia in patients with Parkinson's disease by neural networks. Mov Disord 2000; 15(6): 1104–11.
11. Tu JV, Guerriere MRJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput Biomed Res 1993; 26(3): 220-9.
12. Gueli N, Piccirillo G, Troisi G, Cicconetti P, Meloni F, Ettorre E et al. The influence of lifestyle on cardiovascular risk factors: analysis using a neural network. Arch Gerontol Geriatr 2005; 40(2): 157-72.
13. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002; 35: 352-9.
14. Kattan MW. Comparison of Cox regression with other methods for determining predictive models and normograms. J Urol 2003; 170: 6-10.
15. Bartfay E, Mackillop WJ, Pater JL. Comparing the predictive value of neural network models to logistic regression models on the risk of death for small-cell lung cancer patients. Eur J Cancer Care 2006; 15(2): 115-24.
16. Yamamura S, Takehira R, Kawada K, Nishizawa K, Katayama S, Hirano M et al. Application of artificial neural network modelling to identify severely ill patients whose aminoglycoside concentrations are likely to fall below therapeutic concentrations. J Clin Pharm Ther 2003; 28: 425–32.
17. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons, Inc.; 2000.
18. Rasmussen BK, Jensen R, Schroll M, Olesen J. Epidemiology of headache in a general population-a prevalence study. J Clin Epidemiol 1991; 44: 1147-57.
19. Rasmussen BK. Epidemiology of headache. Cephalalgia 1995; 15: 45-6.
20. Prencipe M, Casini AR, Ferretti C, Santini M, Pezzella F, Scaldaferri N et al. Prevalence of headache in an elderly population: attack frequency, disability, and use of medication. J Neurol Neurosurg Psychiatry 2001; 70(3): 377-81.
21. Headache Classification Committee of the International Headache Society. The Internal Classification of Headache Disorders. Cephalalgia 2004; 24(Suppl 1): 1-150.
22. Silberstein SD, Lipton RB, Dalessio DJ. Overview, diagnosis, and classification of headache. In: Silberstein SD, Lipton RB, Dalessio DJ, editors. Wolff's Headache and Other Head Pain. New York: Oxford University Press; 2001. pp. 6-26.
23. Manel S, Dias JM, Ormerod SJ. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol Model 1999; 120: 337–47.
24. McGonigal MD, Cole J, Schwab CW, Kauder DR, Rotondo MF, Angood PB. A new approach to probability of survival scoring for trauma quality assurance. J Trauma 1993; 34: 863-8.
25. Ebell MH. Artificial neural networks for predicting failure to survive following in-hospital cardiopulmonary resuscitation. J Fam Pract 1993; 36: 297-303.
26. Kappen HJ, Neijt JP. Advanced ovarian cancer: neural network analysis to predict treatment outcome. Ann Oncol 1994; 4: 31-4.
27. Cucchetti A, Vivarelli M, Heaton ND, Phillips S, Piscaglia F, Bolondi L et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. Gut 2007; 56: 253-8.