ELSEVIER

International Journal of Medical Informatics 57 (2000) 1–9

International Journal of Medical Informatics

www.elsevier.com/locate/ijmedinf

Neural network modeling for surgical decisions on traumatic brain injury patients

Yu-Chuan Li [a,b,*], Li Liu [a,b], Wen-Ta Chiu [c,d], Wen-Shan Jian [e]

[a] *Graduate Institute of Medical Informatics, Taipei Medical College, No. 250, Wu-hsin Road, Taipei 110, Taiwan, ROC*
[b] *Center for Biomedical Informatics, Taipei Medical College, Taipei, Taiwan, ROC*
[c] *Department of Medical, Taipei Medical College, Taipei, Taiwan, ROC*
[d] *Wanfang Hospital, Taipei, Taiwan, ROC*
[e] *School of Healthcare Administration, Taipei Medical College, Taipei, Taiwan, ROC*

Received 7 September 1998; accepted 31 August 1999

Abstract

Computerized medical decision support systems have been a major research topic in recent years. Intelligent computer programs were implemented to aid physicians and other medical professionals in making difficult medical decisions. This report compares three different mathematical models for building a traumatic brain injury (TBI) medical decision support system (MDSS). These models were developed based on a large TBI patient database. This MDSS accepts a set of patient data such as the types of skull fracture, Glasgow Coma Scale (GCS), episode of convulsion and return the chance that a neurosurgeon would recommend an open-skull surgery for this patient. The three mathematical models described in this report including a logistic regression model, a multi-layer perceptron (MLP) neural network and a radial-basis-function (RBF) neural network. From the 12 640 patients selected from the database. A randomly drawn 9480 cases were used as the training group to develop/train our models. The other 3160 cases were in the validation group which we used to evaluate the performance of these models. We used sensitivity, specificity, areas under receiver-operating characteristics (ROC) curve and calibration curves as the indicator of how accurate these models are in predicting a neurosurgeon's decision on open-skull surgery. The results showed that, assuming equal importance of sensitivity and specificity, the logistic regression model had a (sensitivity, specificity) of (73%, 68%), compared to (80%, 80%) from the RBF model and (88%, 80%) from the MLP model. The resultant areas under ROC curve for logistic regression, RBF and MLP neural networks are 0.761, 0.880 and 0.897, respectively ($P < 0.05$). Among these models, the logistic regression has noticeably poorer calibration. This study demonstrated the feasibility of applying neural networks as the mechanism for TBI decision support systems based on clinical databases. The results also suggest that neural networks may be a better solution for complex, non-linear medical decision support systems than conventional statistical techniques such as logistic regression. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

* Corresponding author. Tel.: +886-2-23776730; fax: +886-2-27339049.
E-mail address: jack@tmc.edu.tw (Y.-C. Li)

PII: S1386-5056(99)00054-4

Keywords: Medical decision support system; Computer-assisted medical decision; Decision theory; Intracranial hematoma; Traumatic brain injury; Logistic regression; Neural networks

1. Introduction

Computerized medical decision support systems have been a major research topic in recent years. Knowledge-based computer programs were implemented to aid physicians and other medical professionals in making difficult medical decisions such as the determination of surgery for patients with acute abdominal pain [1], the use of antibiotics on patients with nosocomial infection [2], presurgical determination of the types and range of renal tumors [3] and broad-spectrum diagnosis of internal medicine diseases [4,5]. Mathematical and statistical methods have been used also to develop models for clinical diagnosis and treatment as well as policy applications. The most commonly used methods include discriminant analysis, logistic regression, recursive partitioning and neural networks [6–9]. In clinical applications, these systems can potentially enhance diagnostic accuracy and thus result in better treatment decisions and more appropriate use of health care resources. In health policy applications, these systems can provide a model for setting reasonable reimbursement plans, outcome expectations and health care resource allocation.

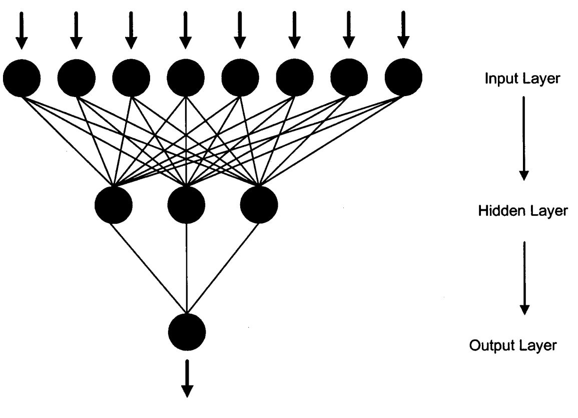


Fig. 1. A multilayer feed-forward neural network.

Among the mathematical and statistical modeling techniques used in medical decision support, neural networks attract much of the attentions in recent studies. Connectionist neural networks are mathematical constructs composed of individual processing elements interacting with each other based on the paradigm of biological neurons. These systems in their most basic implementation consist of a layer of input variables, connected to an intermediate layer of derived variables (a 'hidden' layer), and then to the final output prediction. Processing of multiple events occurs in the hidden layer, with final results passed to the output layer (Fig. 1). The connections between these neurons represent mathematical functions (called 'transfer function' in neural network terms) that propagate the modified 'impulse' to the next neuron. By changing the transfer functions and the parameters associate with these functions, this neural network construct adapt itself to the pattern of the input variables and eventually generates numbers that close to values of the designated output variables [10–15]. Through the hidden layers, neural networks are able to identify relationships among input variables, the structure of which need not to be specified a priori. Although conceptually attractive, neural networks are not analyzed easily based on risks attributable to specific clinical characteristics or statistical significance because a neural network relies on its internal representation of weights and functions to process data instead of straightforward equations like a regression model.

One major advantage of using neural networks in medical decision support systems is that a huge effort of knowledge engineering into the domain knowledge can be saved, provided that sufficient amount of training cases are available. The neural networks train itself without much human intervention. Hence, many large epidemiology databases were analyzed not only by traditional statistical methods, but fed also to neural networks for further insight of the interrelations of the variables. However, the question of whether neural networks can outperform other statistical modeling techniques such as logistic regression has not been settled.

Motor vehicle related traumatic brain injury has been a major public health issue in Taiwan partly because of the prevalent use of motorcycle. Traumatic brain injury accidents often involve intracranial hematoma (ICH) if the impact is large enough to cause bleeding inside the skull. Once an ICH occurs, the decision whether to open the skull to remove the hematoma or wait for the hematoma to be absorbed naturally became a critical decision for the physician. Many variables including medical conditions of the patient, type of trauma and position of the hematoma have to be carefully evaluated before a sound decision can be made. To further investigate the factors involved in such decision, a large patient database for traumatic brain injury (TBI) patient has been established by researchers since 1991. This database collected records of TBI patients from teaching hospitals all over Taiwan. One hundred and thirty-two parameters including sex, age, type of collision, type of fracture, type of neurological deficit and so on were recorded from the medical charts and follow-up visits or calls to the patients [16,17].

Given the rapidly increasing importance of computerized decision aids of clinical models, and particularly given the interest in these models, we sought to examine their respective characteristics and performance by utilizing patient data between 1992 and 1994 in this TBI patient database. We developed two neural network models and a logistic regression model that determine the factors involved in making the decision of open-skull surgery for TBI patients. These models were built using a randomly chosen 75% from the 12 640 cases and the other 25% were used to test the performance of the three models.

2. Materials and methods

This study was conducted using data collected from a nation-wide epidemiological study of traumatic brain injury in Taiwan. Three models, including two different neural networks and a logistic regression, were derived from this data set. The performance of these models was compared using receiver-operating characteristics (ROC) curve area.

2.1. Participating hospitals and data collection

One hundred and sixteen large to medium-sized teaching hospitals with qualified neurosurgical department participated in this study. Patients with injury to skull or face bones, blood vessels, nerves and patient with contusion, concussion with loss of consciousness and intracranial hematoma were included. One hundred and thirty-two parameters ranging from insurance status to Glasgow Coma Scale (GCS) were recorded for each patient.

Thirty-two variables were identified by a senior neurosurgical attending physician as clinically related to the decision of open-skull surgery from the 132 parameters recorded. Computed tomography (CT) scan result, although available as one of the parameters in the database, was intentionally omitted since it has a deterministic effect on the decision of surgery intervention. Result of magnetic reso-

Table 1
Variables in the models

Sex
Age group
Use of helmet
Length of loss of consciousness
Length of amnesia
Presence of amnesia
Episode of convulsion
Presence of neurological deficit
Presence of complications
Presence of cranial fractures

nance image (MRI) or other diagnostic procedures that may have the same effect were not available in this database.

After pruning the cases with missing data in the 32 variables, the number of cases decreased from 18 000 to 12 640. In the second step, a step-wise logistic regression was applied to the remaining data set and 10 variables (see Table 1) were selected as being statistically significant ($P < 0.05$) in predicting of the dependent variable (decision of open-skull surgery).

2.2. Separation of training and validation data sets

From the 12 640 cases, 75% were randomly selected as the training group and the other 25% as the validation group. Cases in the train group ($n = 9480$) were used in the development of the logistic regression and neural network models. The validation group ($n = 3160$) was used to test the performance of these models.

2.3. Modeling methods

Using the training data set, one logistic regression model and two neural network models were developed. They are described in the following paragraphs in detail.

(1) The logistic regression model was constructed using the SPSS for Windows version 6.1 'logistic regression' procedure. No variable selection procedure was applied. All the variables unconditionally entered the logistic regression equation because they were already deemed significant in the second step of the variable selection process.

(2) The first neural network model is a multi-layer perceptron (MLP) neural network. It is a typical feed-forward back-propagation neural network that took the common three-layer topology (Fig. 1). An MLP neural network constructs a decision surface in the data space and tried to discriminate instances with similar features by forming a boundary between them. The MLP neural network in our study was built with an 11-node input layer, a seven-node hidden layer and a one-node output layer. A sigmoid function was chosen to be the transfer function of this neural network. The neural network development software used was Neural Connection 1.0.

(3) The second neural network model is a radial-basis-function (RBF) neural network. It is considered a more recent design of the MLP. The RBF is a supervised feed-forward back-propagation neural network with only one hidden layer. While rather than trying to find a boundary between different classes of instances, it forms clusters in the data space with a 'center' for each cluster. These clusters are then used to classify different groups of data instances. The number of centers and the nonlinear functions used to determine the distance away from each center dictate the performance of a RBF neural net. The RBF neural network in our study used a five-center hidden layer and a spline function as the nonlinear transfer function.

2.4. Statistical methods for model testing

Each of the modeling methods was used to yield a continuous estimate of the chance of having an open-skull surgery. Such a continuous variable estimate provides substantially more information than a simple yes/no prediction. However, for a continuous-scale prediction, using a single cut-off point to measure sensitivity and specificity does not sufficiently describe the performance of the predictive model. Rather, the discriminating power of the model can be better captured by measuring the area under the ROC curve and its calibration, as represented by a comparison of the observed and predicted rates across ranked groupings of the chance estimates.

The ROC curve is a graphical representation of the sensitivity and specificity for a model with continuous predictions, obtained by plotting the observed sensitivity versus 1-specificity. Each point on the curve represents the sensitivity and specificity for a prediction based on classifying a patient as to be operated if their predicted probability of surgery is greater than a particular cut-off value, and the area under the ROC curve is the customary summary statistic of sensitivity and specificity across all cut-off values. In this study, the area under the ROC curve was obtained by plotting sensitivity versus 1-specificity for each possible predictive cut-points and summing the area of the created trapezoids. Statistical differences between areas under ROC curve for different models were evaluated according to the method described by Hanley and McNeil [18–20].

Calibration curves were created by grouping the ordered predicted values into equal size deciles, and then calculating the mean observed response for each of these groups. Such calibration curves show the accuracy of this model prediction across the entire range of risk groups, and also allow detection of particular regions of the prediction scale that are not as accurate as others [21].

3. Results

Of the 9480 training cases, 471 (4.97%) received open-skull surgery, while surgery was performed on 168 (5.32%) of the 3160 validation cases. Patient age ranged from 1 to 99 with an average of 38 years old and the male to female ratio was 67 to 33. Since the validation group and training group were separated randomly, their patient profiles resembled each other.

The calibration curves for the respective models are illustrated in Fig. 2. As shown the logistic regression model tended to over-estimate the chance of surgery, while the two neural network models showed close matches of the observed and predicted probabilities.

The areas under ROC curve on the validation data set for the three models are 0.761, 0.897 and 0.880 for the logistic regression, MLP neural network and RBF neural network models respectively. We did a pair-wise comparison of the areas under ROC curve and found highly significant difference between the logistic regression model and the two neural network models ($P < 0.0001$), but found no statistically significant difference between the RBF and the MLP models ($P > 0.05$). The three ROC curves are shown in Fig. 3.

4. Discussion

A number of studies had been conducted to compare the performance of logistic regression and neural networks in the assessment of risk factors involved in outcome predictions [6,7,22,23]. Some reported better performance of neural networks [6,22], others showed similar or somewhat inferior performance in neural network models compared to their logistic regression counterparts [7,23]. In this study, we used a large TBI patient database to develop models that estimate the chance of neurosurgeons' decision on open-skull surgery using a set of clinical parameters of a patient. The neural network models performed consistently better than the logistic regression model either in terms of the area under ROC curve (Fig. 3) or the calibration curve (Fig. 2).

An optimal pair of sensitivity and specificity can not be determined by the ROC curve along, utilities associate with the outcome of false positive and false negative are also important factors in this decision. However, if we simplified the case and assume a balance importance on sensitivity and specificity, we can see that at certain cut-off point, a sensitivity of 88% and specificity of 80% can be found in the ROC curve of the MLP model. In the RBF model, a sensitivity of 80% and specificity of 80% can be found in the ROC curve. On the other hand, the logistic regression model rendered a sensitivity of 73% and a specificity of 68% in its ROC curve.

The complexity of neural networks does make it difficult to relate their output to input. Hart and Wyatt believe that this 'black box' aspect is a major obstacle to the acceptance of neural networks as one mechanism for the medical decision support systems [24,25]. They argue that, to assess the relevance of a decision aid to a particular patient, the user needs insight into the system's behavior. While it is still debatable whether human experts uses hypothetico-deductive reasoning or 'hunch' more frequently in making a medical diagnosis, an accurate second opinion is often helpful in medical decision making with or without a detailed understanding of how it works.

The neural network models developed in this study provided quite acceptable results given the fact that only ten easily obtainable parameters were used. We intentionally omit the variable that records the result of CT scan since the deterministic effect of this vari-

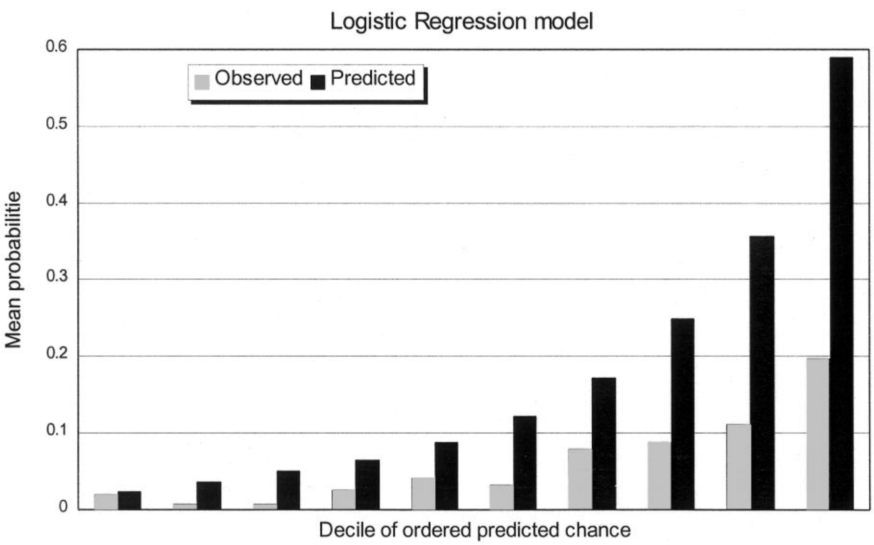
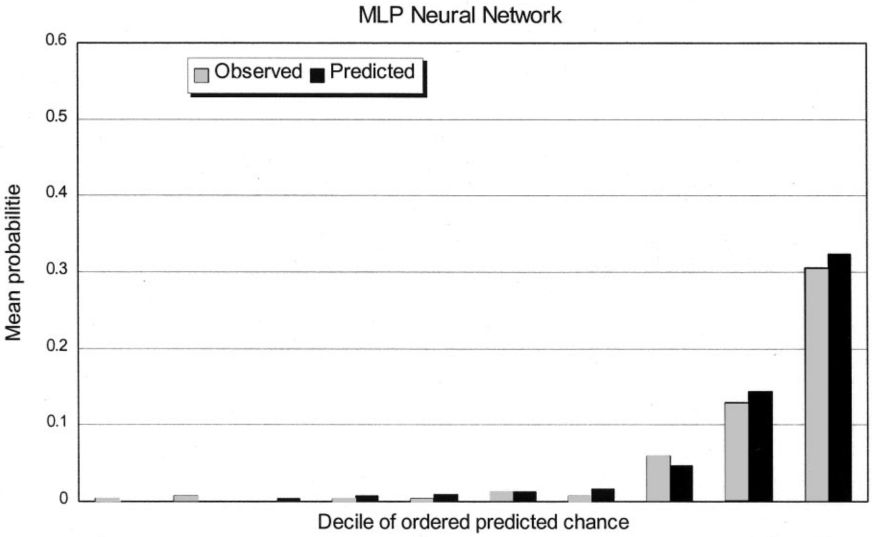


Fig. 2. Predicted and observed mean probabilities of the three models on validation data set ($n = 3160$).



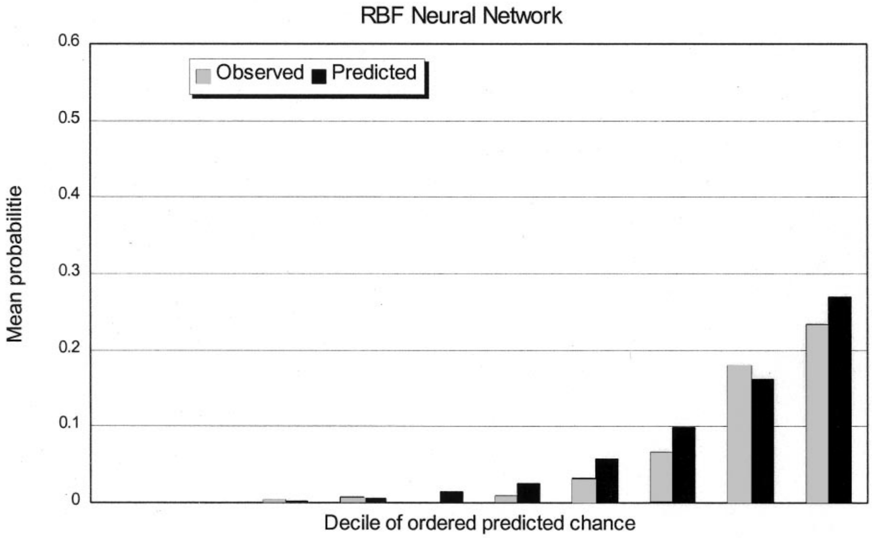


Fig. 2. (*Continued*)

able would wipe out the differences between the tested models. Although the dependent variable represents the decision of surgical intervention by each of the responsible surgeon, in the predictive mode, it can be treated as a reliable estimation of the degree of agreement to operate on a specific patient since the neural network models were trained by a large patient database from teaching hospitals with qualified neurosurgeons. Moreover, omitting this variable also made the resultant model more useful in circumstances where advanced diagnostic tools such as CT scan are not available.

This study demonstrated a significant difference of performance between models based

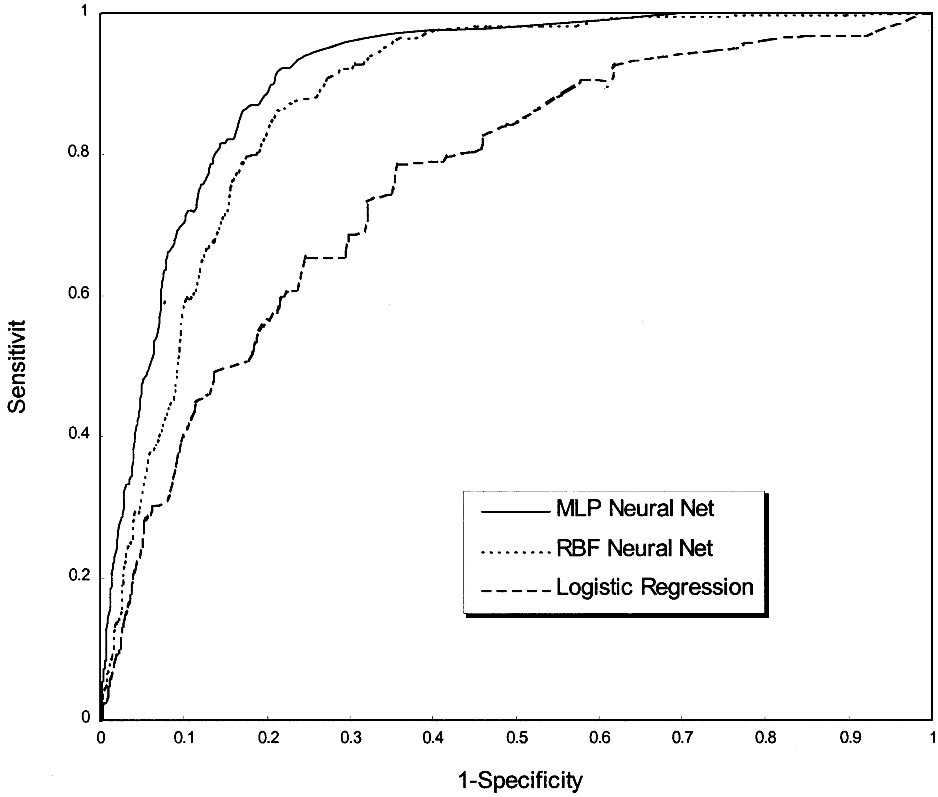


Fig. 3. Receiver-operating characteristic (ROC) curves for the three models on validation data set.

on logistic regression and neural networks to predict the surgical decision for TBI patients. The two neural network models proved better with areas under ROC curve of 0.897 and 0.880, respectively, can potentially be applied to medical decision support systems that advise on the need of open-skull surgery for TBI patients.

Acknowledgements

This work was supported by grant NSC 85-2331-B-038-025 from National Science Council.

References

[1] F.T. de Dombal, Computer-aided decision support — glittering prospects, practical problems, and Pandora's box, Baillieres Clin. Obstet. Gynaecol. 4 (4) (1990) 841–849.
[2] R.S. Evans, S.L. Pestotnik, Applications of medical informatics in antibiotic therapy, Adv. Exp. Med. Biol. 349 (1994) 87–96.
[3] P.L. 1Chang, Y.C. Li, P.J. Huang, T.M. Wang, M.L. Hsieh, The effects of a medical expert system on differential diagnosis of renal masses: a prospective study, Comput. Med. Imaging Graph 20 (1996) 43–48.
[4] Y.C. Li, T. Wong, P.J. Haug, Automated transformation of a large diagnostic knowledge base and the linking to a hospital information system, Abstract from the American Medical Informatics Association Spring Congress, St. Louis, MO, 1993, p. 95.

[5] Y.C. Li, P.J. Haug, Evaluation of the quality of a probabilistic diagnostic system using different inferencing strategies, Proceedings of the 18th Symposium on Computer Applications in Medical Care, Washington DC, 1993, pp. 471–477.
[6] R. Dybowdski, V. Gant, Artificial neural network in pathology and medical laboratories, Lancet 346 (1995) 1203–1207.
[7] H.P Selker, J.L. Griffith, S. Patil, W.J. Long, R.B. D'Agostino, A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients, J. Invest. Med. 43 (5) (1995) 468–476.
[8] B.P. Bergeron, R.S. Shiffman, R.L. Rouse, Data qualification: logic analysis applied toward neural network training, Comput. Biol. Med. 24 (2) (1994) 157–164.
[9] Y. Wu, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, C.E. Metz, Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer, Radiology 187 (1) (1993) 81–87.
[10] W.G. Baxt, The application of the artificial neural network to clinical medicine, Lancet 346 (1995) 1135–1138.
[11] W.G. Baxt, Use of an artificial neural network for the diagnosis of myocardial infarction, Ann. Intern. Med. 115 (1991) 843–848.
[12] Y. Sahasukal, S. Chaithiraphin, P. Panchavinnon, et al., Multivariate analysis in the prediction of death in hospital after acute myocardial infarction, Br. Heart F. 64 (1990) 182–185.
[13] W.G. Baxt, Complexity, Chaos and human physiology: the justification for non-linear neural computational analysis, Cancer Lett. 77 (1994) 85–93.
[14] S.S. Cross, R.F. Harrison, R.L. Kennedy, Introduction to neural networks, Lancet 346 (1995) 1075–1079.
[15] J.M. Barreto, F.M. de Azevedo, Connectionist expert systems as medical decision aid, Artif. Intell. Med. 5 (6) (1993) 515–523.
[16] W.T. Chiu, K.H. Yeh, Y.C. Li, T.H. Gan, H.Y. Chen, C.C. Hung, Traumatic brain injury registry in Taiwan, Neurol. Res. 19 (1997) 261–264.
[17] W.T. Chiu, R.E. LaPorte, T. Songer, D. McCarty, S. Dearcater, Establishment of accurate incidence rates for head and spinal cord injuries in developing and developed countries: a capture–recapture approach, J. Trauma 8 (35) (1993) 206–211.
[18] J.A. Hanley, B.J. McNeil, A method of comparing the area under receiver operating characteristic curves derived from the same cases, Radiology 148 (1983) 839–843.
[19] B.J. McNeil, J.A. Hanley, Statistical approaches to the analysis of receiver operating characteristic (ROC) curves, Med. Decis. Making 4 (1984) 137–150.
[20] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29–36.
[21] J.F. Yates, External correspondence: decomposition of the mean probability score, Org. Behav. Hum. Perform. 30 (1982) 132–156.
[22] T.G. Buchman, K.L. Kubos, A.J. Seidler, M.J. Siegforth, A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit, Crit. Care Med. 22 (1994) 750–762.
[23] M. Nielen, Y.H. Schukken, A. Brand, S. Haring, R.T. Ferwerda-van Zonneveld, Comparison of analysis techniques for on-line detection of clinical mastitis, J. Dairy Sci. 78 (5) (1995) 1050–1061.
[24] J. Wyatt, Nervous about artificial neural networks?, Lancet 346 (1995) 1175–1177.
[25] A. Hart, J. Wyatt, Evaluating black-boxes as medical decision aids, issues arising from a study of neural networks, Med. Inform. 15 (1990) 229–236.