Comparing the Performance of Mathematical Models for Surgical Decisions on Head Injury Patients

Yu-Chuan Li, M.D., Ph.D. [1], Li Liu, Ph.D. [1], Ten-Fang Yang, M.D., Ph.D. [1], Wen-Ta Chiu, M.D., Ph.D. [2]
Center for Biomedical Informatics[1] and Injury Prevention Center[2], Taipei Medical College
Taipei, Taiwan, ROC

This paper compares three mathematical models for surgical decisions on head injury patients. A logistic regression and two neural network models were developed using a large clinical database. Using randomly selected 9480 cases as the training group and another 3160 cases as the validation group. We evaluated the performance of a logistic regression model, a multi-layer perceptron (MLP) neural net and a radial-basis-function (RBF) neural net in terms of their accuracy in predicting physician's decision on open-skull surgery. The resultant area under ROC curve for logistic regression, MLP and RBF neural nets are 0.761, 0.897 and 0.880 respectively. The results 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.

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 also been used 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 [6-7]. 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.

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. The links between these neurons represent mathematical functions (called "transfer function" in neural net terms) that propagate the modified "impulse" to the next neuron. By changing the transfer functions and the parameters associate with these functions, this neural net construct adapt itself to the pattern of the input variables and eventually generates numbers that close to values of the designated output variables [8-9].

One major advantage of using neural nets 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 nets train itself without much human intervention. Hence many large epidemiology databases were analyzed not only by traditional statistical methods, but also fed to neural networks for further insight of the interrelations of the variables. However, the question of whether neural nets can outperform other statistical modeling techniques such as logistic regression has not been settled.
Motor vehicle related head injury has been a major public health issue in Taiwan partly because of the prevalent use of motorcycle. Head 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 head-injury patient has been established by researchers since 1991. This database collected records of head-injury 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 [10-11].

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 head-injury patient database. We developed a logistic regression model and two neural network models that determine the factors involved in making the decision of open-skull surgery. These models were built using a randomly chosen 75% from the 12640 cases and the other 25% were used to test the performance of the three models.

METHODS

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

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 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 Resonance 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 18000 to 12640. In the second step, a step-wise logistic regression was applied to the remaining data set and 11 variables (see Table 1) were selected as being statistically significant ($p < 0.05$) in predicting of the dependent variable (decision of open-skull surgery).

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
Glasgow Coma Scale

Separation of training and validation data sets

From the 12640 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.

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. Each possible value of a categorical variable with more than two values was individually represented in the model construction by a dichotomous “dummy” variable. Replacing the nominally coded variables with dummy variables increased the number of input variables from 11 to 36. All the variables unconditionally entered the logistic regression equation since 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 net. It is a typical feed-forward back-propagation neural net which took the common three-layer topology. 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 net in our study was built with an eleven-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 net. The neural network development software used was Neural Connection 1.0.

3) The second neural network model is a Radial Basis Function (RBF) neural net. It is considered a more recent design of the MLP. The RBF is a supervised feed-forward back-propagation neural net 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 net in our study used a five-center hidden layer and a spline function as the nonlinear transfer function.

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.

Statistical differences between areas under ROC curve for different models were evaluated according to the method described by Hanley and McNeil [12-13].

RESULTS

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

The calibration curves for the respective models are illustrated in Figure 1. 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 net and RBF neural net 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 Figure 2.

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, 14-15]. Some reported better performance of neural networks [6, 14], others showed similar or somewhat inferior performance in neural network models compared to their logistic regression counterparts [7, 15]. In this study, we used a large head-injury 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 (Figure 2) or the calibration curve (Figure 1).

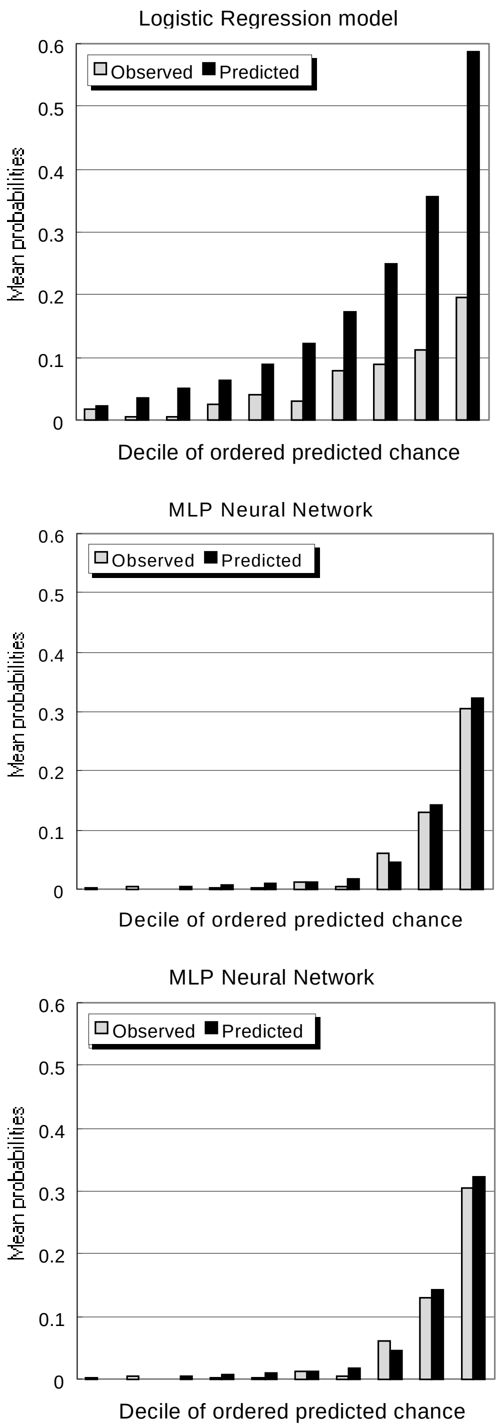

Figure 1. Predicted and observed mean probabilites of the three models on validation data set (n=3160).

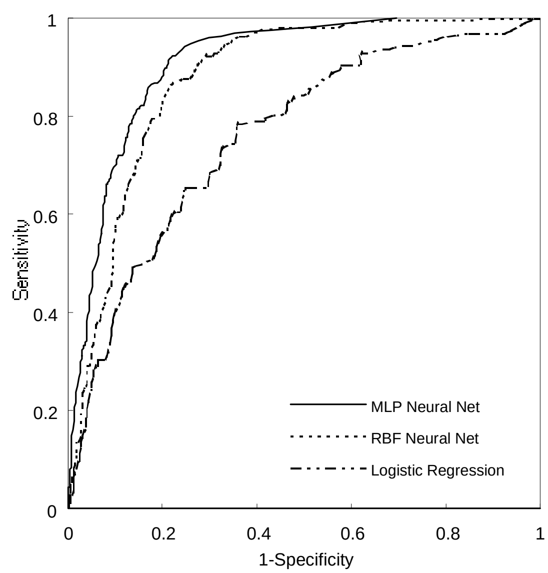

Figure 2. Receiver Operating Characteristic (ROC) curves for the three models on validation data set.

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 cutoff 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 nets 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 nets as one mechanism for the medical decision support systems [16]. 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 net models developed in this study provides quite acceptable results given the fact that only eleven easily obtainable parameters were used. We intentionally omit the variable that records the result of CT scan since the deterministic effect of this variable 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 net 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 on logistic regression and neural networks to predict the surgical decision for head injury patients. The two neural net 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 head injury patients.

Acknowledgments

We thank Wen-Jay Ji for his great help on data reorganization and subsequent analyses. This work was support by grant NSC 85-2331-B-038-025 from National Science Council.

References

1. de Dombal FT. Computer-aided decision support--glittering prospects, practical problems, and Pandora's box. Baillieres Clinical Obstetrics & Gynaecology. 4(4):841-9, 1990.
2. Evans RS, Pestotnik SL. Applications of medical informatics in antibiotic therapy. Advances in Experimental Medicine & Biology. 349:87-96, 1994.
3. Chang PL, Li YC, Huang, Wang TM and Hsieh ML. The effects of a medical expert system on differential diagnosis of renal masses: A prospective study. Comput Med Imag Graph, 1996 (20):43-48.
4. Li YC, Wong T, Haug PJ. Automated transformation of a large diagnostic knowledge base and the linking to a hospital information system. Abstract of American Medical Informatics Association Spring Congress, St. Louis, Missouri, 1993:95.
5. Li YC, Haug PJ. 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, 471-77, 1993.
6. Dybowdski r, Gant V. Artificial nerual network in pathology and medical laboratories Lancet 195; 346: 1203-07.
7. Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB. A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. J Investig Med, 1995;43(5): 468-76.
8. Baxt WG. The application of the artificial neural network to clinical medicine. Lancet 1995; 346:1135-38.
9. Barreto JM, de Azevedo FM. Connectionist expert systems as medical decision aid. Artificial Intelligence in Medicine. 1993;5(6):515-23.
10. Hung CC, Chiu WT, Tsai JC, LaPorte RE, Shih CJ. An Epidemiological study of head injury in Hualien County, Taiwan. J Formosan Med Assoc 1992:1227-33.
11. Chiu WT, LaPorte RE, Songer T, McCarty D, Dearcater S. Establishment of accurate incidence rates for head and spinal cord injuries in developing and developed countries: A capture-recapture approach. J of trauma 1993; 8(35):206-11.
12. Hanley JA, McNeil BJ. A method of comparing the area under receiver operating characteristic curves derived from the same cases.. Radiology 1983; 148: 839-43.
13. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making 1984; 4: 137-50.
14. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ. A comprison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. Crit Care Med 1994; 22:750-62.
15. Nielen M, Schukken YH, Brand A, Haring S, Ferwerda-van Zonneveld RT. Comparison of analysis techniques for on-line detection of clinical mastitis. J Dairy Sci, 1995; 78(5):1050-61.
16. Wyatt J. Nervous about artificial neural networks? Lancet 1995; 346:1175-77.