Use of a Neural Network as a Predictive Instrument for Length of Stay in the Intensive Care Unit Following Cardiac Surgery

Jack V. Tu and Michael R.J. Guerriere

Information Systems Department and Division of General Internal Medicine,
St. Michael's Hospital, University of Toronto, Toronto, Ontario

ABSTRACT

A patient's intensive care unit (ICU) length of stay following cardiac surgery is an important issue in Canada, where cardiovascular intensive care resources are limited and waiting lists for cardiac surgery exist. A predictive instrument for ICU length of stay could lead to improved utilization of existing ICU and operating room resources through better scheduling of patients and staff. We trained a neural network with a database of 713 patients and 15 input variables to predict patients who would have a prolonged ICU length of stay, which we defined as a stay greater than 2 days. In an independent test set of 696 patients, the network was able to stratify patients into three risk groups for prolonged stay (low, intermediate, and high), corresponding to frequencies of prolonged stay of 16.3%, 35.3%, and 60.8% respectively. The performance of the network was also evaluated by calculating the area under the Receiver Operating Characteristic (ROC) curve in the training set, 0.7094 (SE 0.0224), and in the test set, 0.6960 (SE 0.0227). We believe the trained network would be a useful predictive instrument for optimizing the scheduling of cardiac surgery patients in times of limited ICU resources. Neural networks are a new alternative method for developing predictive instruments that offer both advantages and disadvantages when compared to other more widely used statistical techniques.

Key words: Neural networks, intensive care units, cardiac surgery.

INTRODUCTION

In Canada, waiting lists exist for cardiac surgery [1]. Most patients wait several weeks to months before having their cardiac operations. One of the major capacity limiting factors is a shortage of intensive care unit beds [2]. This often results in operations being postponed or cancelled. Simply increasing the number of ICU beds is not viewed as an optimal solution, given the high costs of intensive care and limited amount of health care resources. An alternative approach would be to use existing operating room (OR) and ICU resources in a more efficient manner. If it were possible to predict preoperatively, how long a patient stays in the ICU postoperatively, then patients could be scheduled so as to maximize the utilization of existing operating rooms and ICU beds. For example, patients with longer expected ICU stays could be scheduled in the latter part of the week in order to maximize bed utilization in the ICU over the weekend. This approach would minimize OR and ICU idle time and would lead to the greatest possible number of operations being completed with benefits for patients and staff. The purpose of our study was to develop a predictive instrument for predicting ICU length of stay following cardiac surgery.

Predictive instruments can be developed using a variety of techniques including clinical judgement and statistical modelling [3,4]. A newer approach that has not been widely used in clinical medicine is that of neural networks [5]. Neural networks are computer programs for nonlinear statistical modelling that can be used to perform complex pattern recognition tasks. They were developed by researchers attempting to model the learning processes of the human brain and have successfully been applied in diagnosing acute myocardial

infarction and interpreting chest radiographs among other applications [6,7]. Neural networks have the ability to implicitly identify mathematical relationships between a series of inputs (independent or predictor variables) and the corresponding outputs (dependant or response variables). This is achieved by training the network with a series of training cases consisting of actual numerical inputs and the corresponding outputs. A trained network can then be used to predict the output in an independent test data set. We hypothesized that a neural network might be a useful predictive instrument for preoperatively identifying patients at increased risk for prolonged ICU length of stay following cardiac surgery.

A predictive instrument for predicting ICU length of stay following cardiac surgery has not previously been developed. A number of investigators have identified a group of risk factors for prolonged hospitalization following cardiac surgery. These have included patient age, sex, New York Heart Association (NYHA) functional class, type of surgery, left ventricular function, urgency of surgery, presence of comorbid diseases, and the development of major perioperative complications [8-10]. Since we were interested in predicting a patient's ICU length of stay preoperatively, we restricted our predictor variables to those patient characteristics that would be readily available preoperatively, even though the development of major perioperative complications is clearly an important contributor to overall ICU length of stay.

METHODS

The data set used for developing and testing the neural network was assembled using the Paradox 3.5 (® 1990 by Borland International) relational database program. The study population consisted of all 1409 patients who underwent open heart surgery at St. Michael's Hospital, Toronto, Canada, between Jan. 1, 1990 and Dec. 30, 1991. The data set was divided into two parts: 1) A training set for developing the neural network model consisting of all 713 patients who had cardiac surgery in 1990 and, 2) An independent test set for validating the network consisting of all 696 patients who had cardiac surgery in 1991. St. Michael's Hospital is one of several high risk cardiovascular referral centers for the province of Ontario. There were 4 full-time and 2 part-time cardiac surgeons operating during the study period. The cardiovascular surgery service consisted of an average of 2 operating

rooms, 8 intensive care unit beds, 4 step down beds, and 28 ward beds. Most of the predictor variable information and corresponding ICU lengths of stay were retrieved from the Cardiac Care System (CCS), a computer program used for managing cardiac surgery waiting lists at St. Michael's Hospital. Comorbid disease information was retrieved from the Hospital Medical Records Institute (HMRI) database. Missing information was collected by manual chart review.

Neural network

The neural network model was developed using the Neuroshell 4.1 (® 1991 by Ward Systems Group, Inc.) neural network program. The network uses the standard "back-propagation" algorithm developed by Rumelhart and colleagues. The mathematical equations used in this learning algorithm have been described elsewhere [11]. An IBM-compatible 386Sx computer running at 16 MHz was used to develop the model.
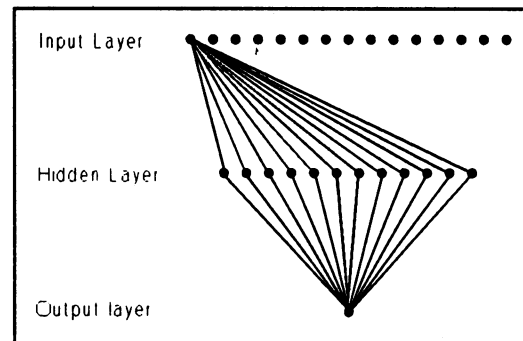

Figure 1. Structure of the neural network. Each circle represents a processing unit and each line represents a connection weight. Only the connection weights for the first processing unit in the input layer are shown.

The neural network is made up of 3 layers of processing units (input, hidden, output) as shown in Figure 1. A series of connection weights links each processing unit in the input layer to each processing unit in the hidden layer, and another series of connection weights links each processing unit in the hidden layer to the processing unit in the output layer. The optimal structure (i.e. number of processing units/layer) and training parameters in the network are discovered by trial and error as there is no method available for prospectively determining them.

The processing units in the input layer are used to encode the values of the input (predictor) variables shown in Table 1. These were variables thought to be possible risk factors for prolonged ICU length of stay. Most of the variables were coded as binary inputs (0 or 1), depending on the absence or presence of the risk factor. Left ventricular function (1-4) was treated as an ordinal variable. Urgency of surgery was coded as 1 for elective operations, 2 for urgent operations (surgery required within 72 hours), and 3 for emergency operations (surgery required within 24 hours). Age was represented as an ordinal variable (1 - Age < 60, 2 - Age 60-70, 3 - Age > 70) after identifying appropriate stratification points using univariate statistical analysis. Type of surgery was divided into 3 separate input units: coronary artery bypass grafting (CABG) surgery or atrial septal defect (ASD) repair alone, single valve surgery alone (i.e. aortic or mitral valve surgery), and all other types of surgery (i.e. combined CABG/valve surgery). The neural network automatically scaled the non-binary input variables into a range (0-1) that could be used by the network.

Patient age group
  - <60 (1), 60-70 (2), >70 (3)
Sex
Left ventricular function (Grade 1-4)
Previous cardiac surgery
Urgency of surgery
  - Elective (1), Urgent (2), Emergency (3)
Type of surgery
  - CABG or ASD surgery alone
  - Single valve surgery (i.e. Aortic or mitral valve replacement)
  - Other surgery (i.e. combined CABG/valve surgery)
Comorbid diseases
  - Hypertension
  - Diabetes
  - Chronic obstructive lung disease
  - Chronic renal failure
  - Previous myocardial infarction
  - Previous cerebrovascular accident
  - Congestive Heart failure

Table 1. Variables used as inputs to the network.

The neural network is trained by an iterative process involving repeated case presentations from the training data set [12]. For each training case, a numerical pattern of input variables is presented to the network at the input layer, and then these values are propagated through the network from the input through the hidden to the output layer. Each input value is multiplied by connection weights, which are initially set randomly by the network. The processing units in the hidden and output layers each sum the weighted inputs to it (input value * connection weight) and then pass the result through a nonlinear logistic function. The result from the nonlinear function becomes the output of that processing unit. Output from the processing unit in the output layer becomes the predicted output of the network. The root mean square (RMS) error is determined by comparing the predicted output of the network with the known (or true) output for each training case. A "back-propagation" algorithm is used by the network to adjust the connection weights linking the processing units in order to minimize the RMS error difference between the known output and the predicted output of the network. The "learning rate" and "momentum" terms control the rate of adjustment of the connection weights and are set by the network developer. Training cases are randomly presented from the training set until the RMS error of the network has converged to a minimum value.
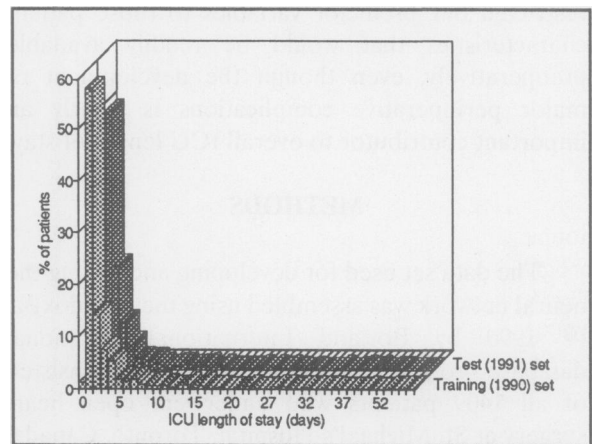

Figure 2. Frequency distribution of ICU length of stay in the training and test data sets.

The frequency distribution of ICU length of stay in the data sets (training and test sets) is shown in Figure 2. ICU length of stay was rounded to the nearest day. We defined a long ICU stay as any stay greater than 2 days and a short stay as less than or equal to 2 days. The stratification point of 2 days was chosen as this was felt to be the most useful outcome measure for scheduling purposes.

668

ICU length of stay is represented in the output processing unit during training as a dichotomous variable with 1 representing long ICU stays and 0 representing short ICU stays. Thus, the output of the trained network becomes a range of values between 0 and 1, with values closer to 0 indicating a greater likelihood of a short stay and values closer to 1 indicating a greater likelihood of a long stay. Although the output from the network is not a probability, it can be viewed and analyzed in a similar manner. Patients with multiple ICU readmissions during one hospital stay had the cumulative ICU stay counted as one prolonged stay rather than as multiple stays. Patients who died in the immediate perioperative period or in the ICU were included in the data sets, as these patients also consume varying amounts of ICU resources, although they represent a different spectrum of risk.

RESULTS

The optimal structure of the neural network model was determined by "trial and error" and was found to be that shown in Figure 1, with 15 processing units in the input layer, 12 processing units in the hidden layer, and 1 processing unit in the output layer. The best learning rate was found to be 0.15 and best momentum term 0.10. The RMS error of the network converged to a minimum value of 0.056211 in the training set and 0.056454 in the test set after 28000 training case presentations.

The performance of the network was evaluated by calculating the area under the Receiver Operating Characteristic curve using a nonparametric method with the ROC Analyzer V5.2 program developed by Dr. Robert M. Centor [13]. The area under the ROC Curve was found to be 0.7094 (SE 0.0224) in the training set and 0.6960 (SE 0.0227) in the test set. A perfect predictive instrument would have an area of 1.00 under the ROC curve, and a non-predictive instrument would have an area of 0.50 or less. The performance of our network lies in between these two extremes and is similar to that of other prediction rules [14].

Three separate risk strata were evaluated in the test data set after choosing two threshold output values of 0.25 and 0.50 as shown in Table 2. This allowed the patients in the test set to be stratified into three risk categories (low - $p < =0.25$, intermediate - $0.25 < p < =0.50$, and high - $p > 0.50$), where p is the predicted output value from the trained network. In the group of low risk patients,

who constitute over half the total patient population, the observed risk of prolonged ICU length of stay was only 16.3%, whereas in the high risk group, the risk of a prolonged stay increased significantly to 60.8%. Patients who were classified as being at intermediate risk were found to have an overall risk for a prolonged ICU stay of 35.3%. Thus, the network was able to stratify patients into three fairly distinct risk categories for prolonged ICU length of stay following cardiac surgery.

Risk category predicted	Short stay patients (%)	Long stay patients (%)
Low ($p < =0.25$)	324 (83.7)	63 (16.3)
Intermediate ($0.25 < p < =0.50$)	152 (64.7)	83 (35.3)
High ($p > 0.50$)	29 (39.2)	74 (60.8)
Total	505 (72.6)	191 (27.4)

Table 2. Performance of the trained neural network in the test data set.
(p = predicted output value from the neural network)

DISCUSSION

This study has demonstrated that a neural network can be used as a predictive instrument to stratify patients into three distinct risk categories for prolonged ICU length of stay following cardiac surgery. The risk stratification scheme that was developed could potentially be applied clinically in situations where only 1 ICU bed is available for several days and two patients equally require surgery. If one patient was in a low risk group for long ICU length of stay and the other was in the high risk group, then scheduling the low risk patient for surgery first would result in both operations being completed within 48 hours (83.7% of the time), whereas operating on the high risk patient first would result in a much lower probability (39.2%) of both operations being completed within a 48 hour time period. The trained network could also be used to identify more "low risk" patients for surgery during times of ICU bed shortages, whereas more "high risk" patients could be operated on when ICU resources are more available, such as prior to a weekend.

The network only partially predicts the variation seen in length of stay as shown by the ROC curve analysis. This is likely a function of the clinical problem under study. ICU length of stay is potentially influenced by multiple factors in addition to preoperative patient characteristics. Patients thought to be at low risk for long stays may develop one of several severe perioperative complications that would have been completely unexpected preoperatively. We did not include the patient's surgeon in our network model because there are no intersurgeon differences in ICU length of stay at our hospital and because it would make the model ungeneralizable. However, other investigators have discovered that there is variation among surgeons and institutions in patient mortality associated with cardiac surgery, even after adjusting for differences in patient case-mix [15]. There is also interinstitution variation in ICU length of stay following cardiac surgery that may make our trained neural network model less generalizable [16].

Whether our neural network model will be accepted by physicians for use clinically remains to be determined. The network could relatively easily be linked to the computer programs currently being used for maintaining cardiac surgery patient waiting lists and a risk stratification category for prolonged ICU length of stay automatically produced. This would represent an additional factor to be considered when scheduling patients, in addition to the patient's need for surgery and risk for perioperative morbidity and mortality. Alternatively, patient characteristics could be entered into the neural network model whenever ICU resources are limited and a need for an ICU length of stay prediction exists.

However, the "black-box" nature of the network may be an important barrier to its acceptance by clinicians. Although the network has learned that patients who are older and have more comorbidity are at an increased risk for prolonged length of stay, it is unable to identify this relationship explicitly. Physicians may be very reluctant to accept the output of a computer program, even when its predictive ability is proven, when there is no explicit explanation for the derivation of the results. They will often however readily accept the recommendations of an expert colleague on faith without questioning the rationale behind their thoughts. Clinicians probably have more faith in the "human neural network" than in an "artificial neural network" because they are comfortable with the

output of the human network and know that it is usually fairly reliable. Educating physicians about the potential usefulness and theory behind artificial neural networks will be necessary before widespread acceptance of this prediction method occurs. Few computerized decision aids have had significant clinical impact, and it remains to be determined whether "artificial neural networks" will be accepted as a useful adjunct to "human neural networks" in medical decision making.

We have demonstrated that a neural network can successfully be used as a predictive instrument for a complex clinical problem. How does it compare as a prediction method for binary outcomes against other more widely used statistical techniques such as logistic regression, recursive partitioning, and discriminant analysis? We believe that neural networks are unlikely to replace these other methods as the prediction method of choice unless they are easier to develop and perform more accurately. Table 3 is a comparison of what we consider to be the major advantages and disadvantages of using neural networks versus logistic regression, the most commonly used statistical method for predicting dichotomous outcomes. Commercially available neural network packages can be used to develop working models by newcomers to neurocomputing within a relatively short time period, when compared with the time required to learn logistic regression techniques or extract rules from experts for expert system development. There is a tremendous amount of research and excitement surrounding the field of neural networks, and new improved learning algorithms are continually being developed by researchers in the field.

At present, the only method for determining whether a neural network is better for a given problem is to compare its performance with that of other statistical methods using the same training and test data sets. Neural networks are unlikely to offer any advantage over regression analysis for problems with relatively simple, linear relationships between independent and dependant variables. However, in complex pattern recognition problems such as voice recognition or image processing, neural networks may significantly outperform conventional regression techniques [17]. The overall utility of neural networks as a predictive instrument in clinical medicine remains to be seen. Further studies, involving a wide variety of data sets, will be required.

Neural networks	Logistic regression
Better for identifying complex, nonlinear relationships between independent and dependent variables.	Better if simple, linear relationships exist between independent and dependent variables.
Developer does not require extensive background in neural network theory.	Developer requires substantial statistical background.
"Black-box" - Relationships and interactions between input variables and predicted outputs cannot be explained.	Relationships between input variables and predicted outputs explicitly identified.
Requires significant computing power and longer computational times.	Model can be developed with much less computational time.
Can handle fuzzy and missing data.	Assumes data is complete and accurate.
Can detect interactions between all input variables implicitly.	Requires explicit interaction terms to be defined.
Model can only be used by those with the trained neural network software.	Model can easily be used by anyone with a hand calculator.
Acceptance as a predictive instrument by clinicians unknown.	Models more readily accepted and familiar to clinicians.

Table 3. Comparison of neural networks and logistic regression as tools for developing a predictive instrument.

In conclusion, we have successfully trained a neural network to predict a patient's risk for a prolonged ICU length of stay following cardiac surgery. The network was able to stratify patients into three heterogenous groups at low, intermediate, and high risk for prolonged ICU stay. The network performed well in an independent test set but awaits prospective clinical testing. We believe our neural network would be a clinically useful predictive instrument for scheduling cardiac surgery operations, but it remains to be determined whether clinicians will be willing to utilize our model as a prediction aid. Neural networks offer a new alternative method to clinical judgement and statistical modelling for developing predictive instruments. Predictive instruments developed from neural networks will need to be subjected to the same rigorous evaluation and field trials as those derived from more conventional methods.

REFERENCES

[1] Naylor CD. A different view of queues in Ontario. Health Affairs. Fall 1991:110-28.

[2] Kaminski VL, Sibbald WJ, Davis EM. Investigation of cardiac surgery at St. Michael's Hospital, Toronto, Ontario: Final report. Feb 15, 1989.

[3] Gustafson DH, Fryback DG, Rose JH, Yick V, Prokop CT, Detmer DE, et al. A decision theory methodology for severity index development. Med Decis Making 1986;6:27-35.

[4] Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med 1985;313:793-9.

[5] Guerriere MRJ, Detsky AS. Neural networks: What are they? Ann Intern Med 1991;115:906-7.

[6] Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991;115:843-8.

[7] Gross GW, Boone JM, Greco-Hunt V, Greenberg B. Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs. Invest Radiol 1990;25:1017-23.

[8] Weintraub WS, Jones EL, Craver J, Guyton R, Cohen C. Determinants of prolonged length of hospital stay after coronary bypass surgery. Circulation 1989;80:276-84.

[9] Katz NM, Ahmed SW, Clark BK, Wallace RB. Predictors of length of hospitalization after cardiac surgery. Ann Thorac Surg 1988;45:656-60.

[10] Lazar HL, Wilcox K, McCormick JR, Roberts AJ. Determinants of discharge following coronary artery bypass graft surgery. Chest 1987;92:800-3.

[11] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533-6.

[12] Sittig DF, Orr JA. Evaluation of a parallel implementation of the learning portion of the backward error propagation neural network: Experiments in artifact identification. In: Clayton PD [ed.] Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care, Washington, D.C. McGraw-Hill. 1991;290-4.

[13] Centor RM, Keightley GE. Receiver operating characteristic (ROC) curve area analysis using the ROC ANALYZER. In: Kingsland LC [ed.] Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care, Washington, D.C. IEEE Computer Society Press. 1989;222-6.

[14] Gilpin EA, Olshen RA, Chatterjee K, Kjekshus J, Moss AJ, Henning H, et al. Predicting 1-year outcome following acute myocardial infarction: Physicians versus computers. Comput Biomed Res 1990;23:46-63.

[15] O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. JAMA 1991;266:803-9.

[16] Mazer CD, Byrick RJ, Sibbald WJ, Chovaz PM, Goodman SJL, Girotti MJ, et al. Post-operative utilization of critical care services by cardiac surgery: A multi-center study in the Canadian health care system. (in press)

[17] Waibel A, Hampshire J. Building blocks for speech: modular neural networks are a new approach to high-performance speech recognition. Byte 1989;14:235-42.