Use of Neural Networks for Prediction of Graft Failure following Liver Transplantation.

Sherri Matis PhD, Howard Doyle MD*, Ignazio Marino MD*,
Richard Mural PhD+ and Edward Uberbacher PhD.

Informatics Group, Computer Science and Mathematics Division,
+Biology Division, Oak Ridge National Laboratory,
Oak Ridge, TN 37831-6364 and *Pittsburgh Transplantation
Institute, Pittsburgh, PA 15213.

Abstract

Liver transplantation is a well-established therapeutic option for patients with end-stage liver disease. However, up to 20% of transplanted livers fail to have adequate function initially, and at least half of those will eventually fail. Accurate, early prediction of outcome may ameliorate this situation by encouraging retransplantation before the patient's condition becomes irreversible.

In this study, clinical information was gathered prospectively for 295 patients who underwent liver transplantation at the University of Pittsburgh Medical Center, and was divided into sets. The feed-forward, fully connected, neural networks had 7 or 8 inputs, a single hidden layer consisting of 3 nodes and a single output node (failure=1, success=0). The networks were trained with data from a randomly selected subset of 240 patients while the remaining 55 patients made up the test set. The preoperative (day 0) data consisted of patient demographics plus the results of standard liver function tests. The "day 1" data consisted information gathered during surgery plus the prediction of outcome from day 0. Data for days 2-5 included results from standard liver function tests plus the prediction of outcome from the previous day's network. The network was trained using a standard back propagation algorithm. Training was assessed by testing the ability of the network to correctly predict the outcome of the 55 patients in the test set. The accuracy of prediction by the neural network improved each day and so by day 5, 98% of the graft survivors in the test set were correctly predicted while 88% of graft failures in the test set were correctly predicted.

Introduction

Outcome prediction is becoming increasingly more important in the management of patient care. There are many cases in which reliable, early determination of prognosis can greatly benefit not only the patient but also allow more efficient use of hospital resources. Computer based medical systems have been developed to accomplish this task[1-8]. By providing an unbiased decision, they could potentially address many legal and ethical issues surrounding patient treatment. A speedy and accurate prediction of outcome would afford medical personal more time to prepare, for example, for retransplantation of an organ because of failure of the primary transplant. Finally, more reliable prediction strategies will, in many cases, increase the patient's well being by providing a positive answer about the outcome of a procedure.

Liver transplantation is now a well established therapeutic option for many patients with end stage liver disease [2,5,9]. There are many patients who are candidates for transplantation, but only a limited supply of donor organs. There is considerable patient pathology associated with the terminal stages of liver disease, and in as many as 20% of the cases the donor organs fail early in the post-transplant period. Even if a substitute donor organ is found for retransplantation, there is a significant morbidity associated with liver failure. In order to more efficiently manage the patient population prior to surgery and following it, we need a systemic method of early, reliable, prediction of patient outcomes following these procedures.

Methods designed to predict patient prognosis are not new to science. Scoring systems which use clinical test data, such as APACHE III, can be used to accurately assess the medical status of patients[8]. However, for certain subclasses of patients, such as those with multiple organ failure or liver transplant patients, the Apache score is not predictive[10]. Child-Pugh scores can be calculated based on liver function [11] but these have been shown to have little predictive value for liver transplantation[12]. Other schemes for prediction of the progress of biliary cirrhosis utilize the results of invasive procedures in the formulation of a statistically based disease model [13,14]. A linear regression model has been developed for liver transplantation but the predictive value of the model is still poor possibly reflecting the complexity of the pathologic process or the fact that the modeling protocol is poorly suited to the task[15].

A new wave in medicine uses artificial neural networks for classification problems where a decision is based on the recognition of complex patterns within the data. Artificial neural networks have been shown to make accurate decisions in numerous medical conditions, such as myocardial infarction[1], hepatocellular carcinoma[2] and breast cancer staging[3]. A previously reported study which used neural networks trained on pre-operative and very early post-operative data for the prediction of outcomes following liver transplantation gave promising results[5]. We will describe a somewhat different approach which uses multiple neural networks which are trained on both pre- and post-transplant data. Our system uses the clinical data for each day plus the prediction of the neural network trained with the previous day's data to

|

predict the likelihood of liver failure following liver transplantation.

Methods

Patient population.

Pre- and post-operative data was prospectively collected from 295 adult patients who underwent liver transplantation at the University of Pittsburgh Medical Center, during the years 1991-93. For the purposes of our study, graft failure is defined as patient death or retransplantation within 90 days of surgery, in patients that survived for at least 24 hours after surgery.

Data from both primary transplants and 51 re-transplants was included in the dataset and were designated as such. In some cases, the organ failed prior to day 5 and for these cases, the data for the patient was removed from the test or training set for the days following organ failure.

These data were obtained from an observational study, where patients were treated according to established protocols, so Institutional Review Board approval was unnecessary. In order to maintain confidentiality, the data was anonymous, meaning that no patient identifiers were included with the patient records.

Patient variables.

The data used for system development was obtained from a clinical database maintained at the Pittsburgh Transplantation Institute. Day 0 values where taken on the day prior to surgery. Day 1 values include data which were taken during surgery. Values for days 2-5 were taken on the subsequent days. Table 1 shows all data parameters used for neural network training.

There are essentially two classes of variables used in this study: binary and graded variables. Binary variables consist of a yes or no answer, for our purposes 0 for no and 1 for yes. These data correspond to parameters such as whether the transplant is a primary or retransplant or whether the patient was in the ICU prior to the transplant. Graded data values were normalized to vary from 0 to 1. An example of these types of data are values for serum lactate. The method for normalization is described in the table.

In a few cases, some data were missing from the patient record. In most of these cases, the surgery was clearly a success and in these situations some laboratory tests were not done. In patients who are in obvious decline, it is unnecessary to run certain additional tests. In some cases, the data are simply unavailable in the database. The majority of missing data were day 5 values for serum lactate in patients who showed long term survival. It has been found in earlier studies that serum lactate has low discriminating power[15]. In these cases where a patient's test results were unobtainable, the previous day's value was substituted. In cases where test data is unavailable and values were available for the same test on the previous and following days, the missing value was interpolated. In the case of missing binary data or missing

intraoperative values, the patient was eliminated from the study.

Table 1. Patient parameters used in network training.

Variables	Description	Day[1]	Variable Type[3]
age	Patient age	0	graded
redo	Whether the transplant in question was a retransplantation (1) or a primary transplant (0)	0	binary[1]
pre_mv	Whether the patient was on a mechanical ventilator	0	binary
in_icu	The patient was (1) or was not (0) in the ICU prior to the transplant	0	binary
bili_o	Preoperative bilirubin	0	graded
pt_o	Preoperative prothrombin time	0	graded
cr_o	Preoperative creatinine	0	graded
peak_lac	Peak intraoperative serum lactate	1	graded
bili1-bili5	Postoperative bilirubin (post-op days 1 to 5)	1-5	graded
sgot1-sgot5	Postoperative aspartate aminotransferase	1-5	graded
lac1-lac5	Postoperative serum lactate	1-5	graded
cr1-lac5	Postoperative serum creatinine	1-5	graded
pt1-pt5	Postoperative prothrombin times	1-5	graded
ffp1-ffp5	Whether the patient received fresh frozen plasma (coagulation factors) that day.	1-5	binary
netout[2]	Output from previous days net	1-5	graded

[1]Day corresponds to the neural network for which the value was used as an input value. A total of 6 neural networks were trained, one for each day.

[2]Binary values were either 0 for no or 1 for yes. Graded values were determined by setting the maximum value for the data parameter equal to 1.0, the minimum value = 0.0, and scaling the values for each variable.

[3]The output of the previous day's net was used as an input for each network from day 1-5.

Neural network architectures.

The networks were created using Neuralworks (Neuralware, Pgh PA). A series of feed-forward, back propagation networks were used, one for each day beginning with day 0. All networks had 3 layers; input, a single hidden, and output layer. All networks for days 1-5 had the output of the previous days network as one of the input parameters. A basic diagram of the network architecture is shown in figure 1.

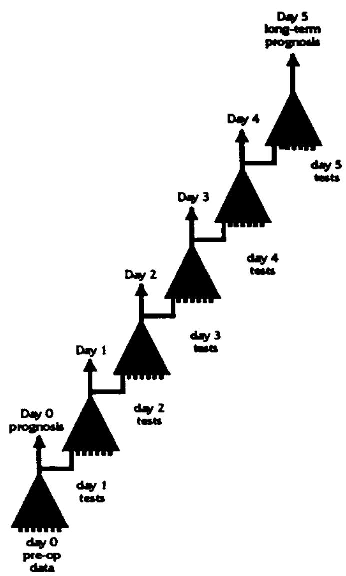

Figure 1. System Architecture.
Organization of the neural network based system used to predict liver failure after transplantation. A separate network was trained for days 0-5. The output from days 0-4 for each patient was used as one of the inputs for the next days network. All neural networks had 7 input nodes, 3 hidden nodes and 1 output node except for day 1 which had 1 additional input node for a total of 8 inputs. The data used to train the networks is described in table 1.

Network training and testing

A population of 295 patients who underwent liver transplantation from 1985-1993 was in this study. In this group, 244 (83%) of the patients were primary transplants while the remaining group of 51 (17%) patients were undergoing a retransplantation. The data set consisted of 44 patients with final outcomes of graft failure (output = 1) and 251 patients with successful grafts (output = 0). Thus, in the dataset, the livers failed in 15% of the patients. Of the 295 patients, 240 were included in the training set and the 55 remaining patients were used for the test set.

The network for day 0 was trained on data from the 240 patients which included 34 cases of liver failure. After training, the training data was processed by the day 0 network to obtain a score which was then used as one of the inputs for the day 1 network. Subsequent nets were tested and trained in the same way and typically a network was trained with about 2500 examples. If a patient rejected the donor liver on or before day 5, their data were removed from the following day's networks. Following the training of each days' network, the network performance was measured by testing the ability of the network to correctly score and thus classify examples in the independent test set.

The values for day 0 were taken on the day prior to surgery and are shown in table 1. The patients age was included along with whether the patient was in the ICU, whether the patient had been previously transplanted and preoperative liver function tests. Since many patients deteriorate rapidly once their liver start to fail, the values for the tests reflect the over all homeostasis of the patient prior to surgery, which influences the outcome following surgery.

Day 1 values were taken on the post-operative day 1 (except for peak interoperative lactate), that is, within 24 hours of the operation. All of these values are indicative of liver function. The outcome prediction from the day 0 network was also input into this network.

Day 2-5 values were the results of liver function tests for that day plus the network prediction from the previous day.

Results

Network prediction

The neural network predictions for each day are shown in table 2. The results are given for the test set of 55 patients. The network output thresholds were chosen to best separate the data accurately into failure and survivor subsets and were typically at about 0.20. Even prior to surgery, using data from day 0 only, the network was 70% accurate at predicting liver failure and 87% accurate at predicting liver survival. For the entire test set on day 0, the accuracy of network prediction was 83%.

The prediction accuracy continued to improve in the following days until on day 5, the prediction accuracy was 88% for liver failure and 98% for liver survival with an overall accuracy of 96%. The extremely low false positive rate of 2% makes this system useful in prediction of liver failure following liver transplantation.

Day	%Pos	%Neg	%Total	Total Pos	Total Neg
0	70%	87%	83%	10	45
1	80%	84%	85%	10	45
2	90%	89%	87%	10	45
3	89%	91%	89%	9	45
4	88%	91%	91%	8	44
5	88%	98%	96%	8	44

Table 2. Results of the neural network's prediction for the test set for each day. As the study progressed, some patients were eliminated from the study due to liver failure and retransplantation or because data were missing and could not be interpolated.

Discussion

Artificial neural networks were first developed as a model for the network of neurons in the brain, but early work showed that they were perhaps too simplistic[16]. They have, however, proved to be very useful for adaptive pattern recognition. The back propagation algorithm used in network training has allowed the networks to learn complex relationships present in data, which influence classification [17].

We have presented the results of a neural network based system designed to predict liver failure in patients following liver transplantation. We used a series of neural networks trained on daily data parameters taken from the day prior to liver transplantation to day 5 following the procedure. Many of these tests, when used independently, have been shown to be poor discriminators of graft outcome [15]. The neural network can integrate many weak components of this complex data and learn how to make accurate decisions. The system accurately predicts 88% of the cases in which the newly transplanted liver failed. For graft survival, it was 98% accurate in it's predictive ability. Although the overall accuracy of the network is quite high, some improvements are possible.

A possible weakness of the neural network based system is that we have utilized only patient parameters in the system design. However, there is a significant contribution to organ survival, from the donor organ itself. The length of time the organ has been perfused, the histocompatibility of the donor tissue, the health of the donor and the case of the donor's death may all be contributing factors in the function of the donor organ following transplantation. Addition of these parameters and subsequent increases in the size of the patient population should increase the discrimination ability of the resulting classification system.

The ability of computational classification systems has many positive implications in medical science, particularly in a surgical setting. Other systems have been developed, for instance, APACHE III which uses a statistical approach (logistic regression) to classify patients in the ICU[8]. Many other neural network based systems have been developed for classification of a diverse group of problems, from image analysis to predicting outbreaks of food-borne illness[1-5,18,19]. These systems can be used to make an unbiased assessment in a clinical setting. While computer based systems are not completely faultless, they are capable of working for many hours and are capable of processing large amounts of data. In one study, the APACHE system outperformed the ICU staff in evaluating the chance of patient survival in a group of critically ill patients[20] although for other problems the APACHE II system is not adequate[10]. These systems are not designed as stand alone systems which will eliminate the need for decisions from the ICU staff or transplant surgeon. They are designed to provide additional information for the decision process. Due to current reforms in the health care industry, there is a move to increase the quality of care while decreasing the cost. In many instances, computer based systems can be used to accomplish this goal.

REFERENCES.

[1] Baxt, W.G. 1991. Ann. Intern. Med. 115:843-848.
[2] Doyle. H. *et al.* in press. Proceedings of the First World Congress on Computational Medicine, Public Health, and Biotechnology.
[3] Burke, H., Goodman, P, Rosen, D. 1994. Proceedings of the IEEE Seventh Symposium on Computer Based Medical Systems. IEEE Computer Society Press.
[4] Uberbacher, E. C. and Doyle, H. 1993. Personal communication.
[5] Doyle, H. et al. 1994. Ann. Surg. 219:408-415.
[6] Buchanan, B. G. and Shortcliffe, E. H. 1985. Rule-Based Expert Systems Addison-Wesley, Reading, Massachusetts.
[7] Knaus, W. A. *et al.* 1985. Crit. Care Med. 1985. 13:818.
[8] Watts, C. M. in press. Proceedings of the First World Congress on Computational Medicine, Public Health, and Biotechnology.
[9] Dickson, R. *et al.* 1989. Hepatology. 10:1.
[10] Cerra, F. B., Negro, F. and Abrams, J. 1990. Arch. Surg. 125:519.
[11] Garden, O.J. *et al.* 1985. Br. J. Surg 72:91.
[12] Oellerich, M. *et al.* 1991. Transplantation 51:801.
[13] Roll, J., Boyer, J. L. and Barry, D. 1983. N. Eng. J. Med. 308:1.
[14] Christensen, E. *et al.* 1985. Gastroenterology, 89:1084.
[15] Doyle, H. *et al.* 1994. Transplantation, 57:1028.
[16] Rosenblatt, F. 1958. Psychol Rev. 65:386.
[17] Rumelhart, D. E., Hinton, G. E., Williams, R. L. 1986, Nature 323:533.
[18] Gross, G. W. *et al.* 1990. Invest. Radiol. 25:1017.
[19] Bianchi, G. *et al.* 1994. in Proceedings of the 7th IEEE Symposium on Computer-Based Medical Systems. IEEE Computer Society Press. Los Amamitos Ca.
[20] Chang, R. W. S. *et al.* 1989. Crit Care Med. 17:1091.