INTELLIGENT RISK ANALYSIS BY SELF-ORGANISING MAPS – THE PROBABILITY OF PATIENT SURVIVAL FOLLOWING LIVER TRANSPLANTATION

Yrjö Hiltunen*, Geoffrey H. Haydon**, Nick Murphy**, Bridget Gunson**, David Collett***, James Neuberger**

* University of Kuopio, Department of Environmental Sciences, Kuopio, Finland
** The Queen Elizabeth Hospital, Liver Unit, Birmingham, United Kingdom
*** UK Transplant, Statistics and Audit Directorate, Bristol, United Kingdom

yrjo.hiltunen@uku.fi

Abstract: Self-organising maps (SOM) have been successfully applied in many fields of research and are thus likely to be a useful tool for medical risk analysis. In this paper, we demonstrate the use of SOM-based analysis for predicting the probability of patient survival following liver transplantation. The SOM was trained using a dataset of recipient, donor and peri-operative factors from 381 patients transplanted for fulminant hepatic failure (FHF) resulting from non-A and non-B hepatitis. Reference vectors of the SOM were then classified into three clusters. Survival probabilities were assigned for these clusters using follow up information on these patients. The results show that the method can be successfully applied to predict post-transplant survival probabilities of patients with fulminant non-A non-B hepatitis.

Introduction

Fulminant hepatic failure (FHF) with non-A and non-B hepatitis accounts for 25% of all cases of acute liver failure admitted to liver units in the UK. Liver transplantation is often the only form of therapy available for these patients. However, graft and patient survival have been demonstrated to differ widely because of a range of recipient, donor and peri-operative factors.

Recent applications have demonstrated that artificial neural networks can provide an efficient and highly automated method for modelling liver transplantation [1-3] or other medical data [4-7]. In particular, clinical studies, which use standardised protocols, are most likely to benefit from automated ANN analysis [4]. Self-organising maps [8] have been successfully applied in many areas of research and are thus also considered a potential tool for medical data analysis. SOMs offer an efficient means of handling complex multidimensional data; for example, donor and recipient factors that might determine outcome in liver transplantation.

In this study we have constructed a self-organising map in a population of patients transplanted for FHF resulting from non-A, non-B hepatitis and then used it to examine survival probabilities and the relationship between donor and recipient factors following liver transplantation.

Materials and Methods

Patients: The dataset consisted of 50 (30 recipient, 19 donor and 1 peri-operative) factors (inputs) from 381 (256 female; median age 40 years) consecutive liver transplants for fulminant non-A non-B hepatitis undertaken between 07/05/1994 and 11/11/2002 from all regional transplant centres in the UK (Table 1). Donor and recipient data were included from the time of offering the organ to the transplant surgeon (t=0).

Self-organising maps: Self-organising maps (SOMs) are an artificial neural network methodology, which can transform an n- dimensional input vector into a one- or two-dimensional discrete map. The input vectors, which have common features, are projected to the same area of the map e.g. (in this case described as "neurons"). Each neuron is associated with an *n*-dimensional reference vector, which provides a link between the output and input spaces. During learning, the input data vector is mapped onto a particular neuron based on the minimal *n*-dimensional distance between the input vector and the reference vectors of the neurons. Then the reference vectors of the activated neurons are updated. When the trained map is applied, the best matching neurons are calculated using these reference vectors. In this unsupervised methodology, the SOM can be constructed without previous *a priori* knowledge [8].

The SOM had 100 neurons in a 10x10 arrangement. All binary input values were coded as 0 and +1 while the continuous inputs were transformed to the range (–1, +1). The SOM Toolbox programme (version 2.0 beta) was used in the analysis under a Matlab-software platform (Mathworks, Natwick, MA, USA).

Clustering method: The K-means method is a non-hierarchical cluster algorithm [9]. The basic version begins by randomly picking *K* cluster centers, assigning each point to the cluster whose mean is closest in a Euclidean distances sense, then computing the mean vectors of the points assigned to each cluster, and using these as new centers in an iterative approach.

ISSN: 1727-1983

Survival probabilities: For each of three clusters, the probability p_i of patient survival at 1 month, 3 months and 12 months was calculated using the equation as follows:

$$p_i = \frac{Number\ of\ \ patients\ in\ cluster\ i\ who\ survived}{Total\ number\ of\ \ patients\ in\ cluster\ i} \quad (1)$$

The small number of clusters was chosen to increase the number of patients in each cluster and thus increase the statistical reliability of the survival probabilities. Statistical significance was evaluated by means of the Chi squared test.

Kaplan-Mayer formula: In the presence of censoring, the survival probabilities for each cluster can be generally estimated using the Kaplan-Meier estimator, whose formula is as follows:

$$\hat{p}_i(t) = \prod_{i|t_{(i)} \le t} \left[1 - \frac{d_i}{r_i} \right] \quad (2)$$

where $t_{(1)} \le t_{(2)} \le \ldots \le t_{(n)}$ are the ordered survival times, r_i is the number of individuals at risk just before $t_{(i)}$ (including those censored at $t_{(i)}$) and d_i is the number of patients who die.

Subtraction analysis of reference vectors: For each neuron the reference vectors, which represent the common features of the recipients and donors in each neuron, are defined during the training of the map. In the subtraction analysis, reference vectors of two neurons are subtracted from each other. This method can be used for identification of any differences in recipient and donor factors between corresponding subpopulations of transplanted patients.

Table 1. The recipient donor and operation factors used in the SOM analysis. The number in parenthesis indicates the number of coded inputs.

Recipient data	Serum sodium (1)
Recipient Demographics	Serum potassium (1)
Age (1)	Anti-HCV positive (1)
Gender (1)	HBsAg positive (1)
Ethnic origin (1)	**Donor data**
Co-Morbidity	**Donor Demographics**
Asthma (1)	Donor age (1)
Inflammatory bowel disease (1)	Donor gender (1)
TIPSS (1)	Ethnic origin (1)
Previous upper abdominal surgery (1)	**Donor Co-morbidity**
Portal hypertension (1)	Mode of donor death (3)
BMI (1)	Donor BMI (1)
Inpatient at time of transplantation (1)	Number of days donor ventilated (1)
Ventilated at time of transplantation (1)	Donor hypotensive (1)
Renal support at time of transplant (1)	Donor supported with inotropes (1)
Ascites (1)	**Laboratory Data**
Diuretics (1)	Blood urea (1)
Encephalopathy (1)	Serum creatinine (1)
History of variceal bleeding (1)	Serum sodium (1)
Laboratory Results	Serum potassium (1)
Haemoglobin (1)	Bilirubin (1)
White blood cell count (1)	Aspartate transaminase (AST) (1)
Platelet count (1)	Alkaline phosphatase (1)
Blood urea (1)	Albumin (1)
Serum creatinine (1)	Type of graft (1)
Albumin (1)	**Operation data**
INR (1)	Cold iscemia time
Bilirubin (1)	

ISSN: 1727-1983

Results

The SOM was obtained by training a self-organising network with the data of 381 patients transplanted for FHF with non-A, non-B hepatitis. Figure 1 shows the map and some reference vectors. The three clusters calculated by the k-means method are also illustrated in Figure 1.

The survival probabilities of these three clusters at 1 month, 3 months and 1 year were calculated using Equation 1. The results are in Table 2. There was a statistical difference in the probability of survival of patients between these three clusters (p-values were 0.0082, 0.0005 and 0.0055 for the probabilities at 1 month, 3 months and 1 year, respectively).

The probabilities were further classified using three different categories: low, intermediate and high probabilities of survival after transplantation. These categories are in Table 2.

Figure 2 shows the survival probabilities for patients in the three clusters estimated by the Kaplan-Meier formula (Equation 2).

Subtraction analysis of the reference vectors 1 and 3a and the vectors 1 and 3b revealed differences in donor and recipient factors between corresponding subpopulations of transplanted patients (Figure 3.). The factors are shown in Figure 3.

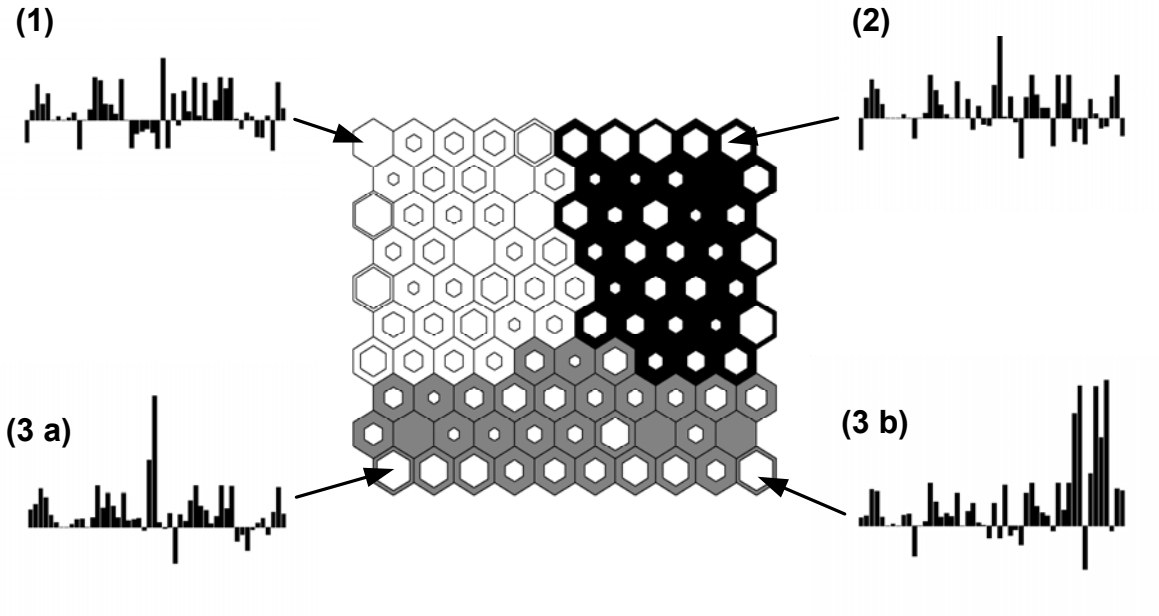

Figure 1. SOM using the data of patients transplanted for FHF with non-A, non-B hepatitis showing the number of the hits on the size of the depicted neuron. The reference vectors related to four example neurons are shown. The background colours show the three main clusters of the map. The inputs 1-30 are from recipient data, 31-49 from donor data and the input 50 from peri-operative data.

Table 2. Patient survival probabilities, number of patients, 95% confidence intervals and user-friendly probabilities for each cluster following liver transplantation determined using SOM method..

Parameter	Cluster 1	Cluster 2	Cluster 3
Probability of Survival at 1 month Number of patients 95% confidence interval	0.836 128 0.760-0.896	0.765 n=117 0.684-0.833	0.667 n=133 0.574-0.751
Probability of Survival at 3 months Number of patients 95% confidence interval	0.803 127 0.723-0.868	0.684 n=117 0.598-0.762	0.573 n=133 0.478-0.664
Probability of Survival at 12 months Number of patients 95% confidence interval	0.730 122 0.642-0.806	0.642 112 0.551-0.727	0.527 123 0.430-0.622
"User friendly "probability	High	Intermediate	Low

ISSN: 1727-1983

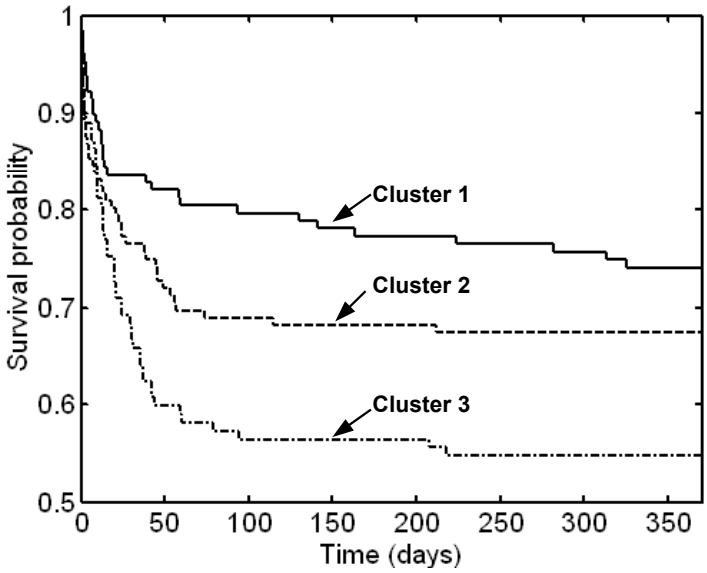

Figure 2. Survival probabilities for each cluster following liver transplantation determined using Kaplan-Meier formula.

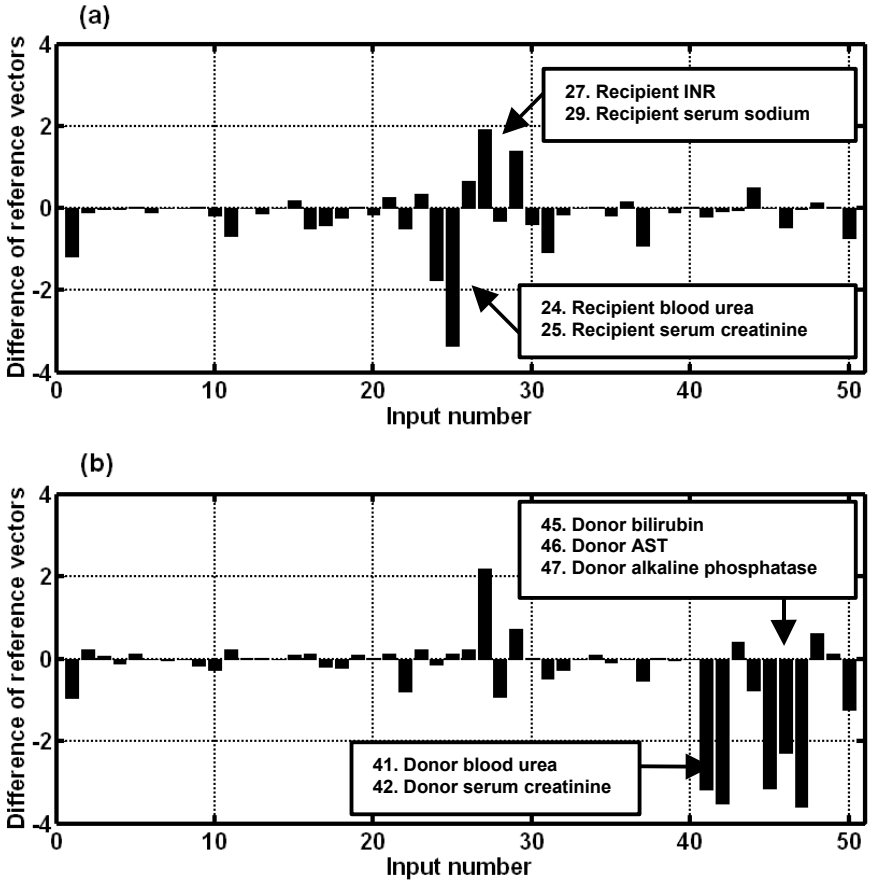

Figure 3. Subtraction analysis of reference vectors reveals the difference in recipient-donor profile between (a) the reference vectors 1 and 3a and between (b) the reference vectors 1 and 3b. The inputs 1-30 are from recipient data, 31-49 from donor data and the input 50 from peri-operation data.

ISSN: 1727-1983

Discussion

The self-organising map described in this study was trained using a data set obtained from all the liver transplant units in the UK. Clusters of the map associated with high, intermediate and low survival are clearly identified, confirming the "user-friendly" nature of the model. The present study confirms a previous report [3], which indicated that both donor and recipient characteristics influence the outcome after liver transplantation in patients with chronic liver disease. Unfortunately, in the present case we did not have enough patients for a validation of the method. However, it has to be emphasized that the follow up information was not used in the training process of the map, which makes the results statistically significant. Haydon et. al. [3] have also shown that this kind of approach provides survival probabilities following liver transplantation for chronic liver disease with a high degree of statistical confidence.

The results of the Kaplan-Meier formula in Figure 2 show that the patients mostly die during the first 60 days post-transplant and after that the probabilities of the clusters 2 and 3 are almost constant. Only the survival probability of cluster 1 decreases after the first 60 days and approaches the probability of cluster 2.

The subtraction analysis, described in this study, can be used for identification of differences in recipient and donor factors between corresponding subpopulations of transplanted patients. Figure 3a shows the difference between the reference vectors 1 and 3a (see Figure 1). The differences are largest with the recipient variables 24 (higher blood urea associated with a lower probability of survival), 25 (higher serum creatinine associated with a lower probability of survival), 27 (higher INR associated with a lower probability of survival) and 29 (lower serum sodium associated with a lower probability of survival), which means that the recipients in subpopulation 3a were more unwell than in the subpopulation 1, but that the donor livers were of similar quality. On the other hand, in the case of the subpopulation 3b, the reason for the low probability of survival post-transplant is likely to be poor quality donor livers utilised (Figure 3b). The difference between reference vectors in terms of donor variables is large (e.g. blood urea, serum creatinine, bilirubin, aspartate transaminase and alkaline phosphatase) and only one recipient variable (INR) differs significantly between the two subpopulations. Therefore the cluster 3 in Figure 3 can be split by the k-means method into two parts: firstly, cluster 3a, where the “user-friendly” survival probability is low and is recipient dependent; secondly, cluster 3b, where the probability of survival is also low, but this time is donor dependent. The complete map is shown in Figure 4.

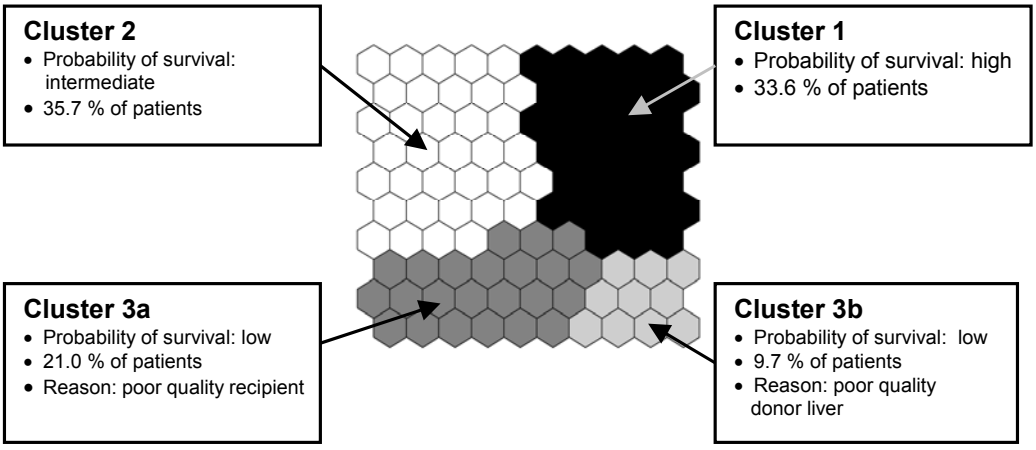

Figure 4. SOM using the data of patients transplanted for FHF with non-A, non-B hepatitis. The background colours show the four clusters of the map. User-friendly probabilities and percentage values of patients for each cluster and reasons of low probability for the clusters 3a and 3b are also shown.

ISSN: 1727-1983

We have successfully applied SOM analysis to predict post-transplant survival of patients with fulminant non-A, non-B hepatitis using variables available before the allocation of a given liver. The trained SOM can be operated using a personal computer or even a PDA device and provides an instant answer with a high degree of confidence when predicting survival of individual patients at three different time intervals following liver transplantation. SOM analysis is likely to have a profound impact when assessing the potential success of liver transplantation in patients with fulminant non-A, non-B hepatitis. For example, in some cases of fulminant non-A, non-B hepatitis, the Transplant Physician and Surgeon has to decide whether to use the scarce resource of a donor liver in an individual whose probability of survival is very low with or without a liver transplant. In these cases, there may be a more deserving recipient of the donor liver, with a predicted higher probability of survival (by SOM analysis).

Conclusions

SOM analysis provides an efficient and automated method for stratifying the probability of survival of individual patients in a population of patients with fulminant non-A non-B hepatitis. The model not only assesses the pre-transplant condition of the patient, but also considers a wide range of donor factors when predicting the probability of survival post-transplant. This unique resource can immediately assess the probability of survival of patients with fulminant non-A non-B hepatitis listed super-urgently for liver transplantation.

References

[1] BANERJEE, R., DAS, A., CHOSHAL, U.C. AND SINHA, M. (2003): 'Predicting mortality in patients with cirrhosis of liver with application of neural network technology', *J. Gastroenterol Hepatol*., **18**, pp. 1054-60

[2] MARSH, J.W., FINKELSTEIN, S.D., DEMETRIS, A.J., SWALSKY, P.A., SASATOMI, E., BANDOS, A., SUBOTIN, M. AND DVORCHIK I. (2003): 'Genotyping of hepatocellular carcinoma in liver transplant recipients adds redictive power for determining recurrence-free survival', *Liver Transpl*., **9**, pp. 664-71

[3] HAYDON, G.H., HILTUNEN, Y., LUCEY, M.R., COLLETT, D., GUNSON, B., MURPHY, N., NIGHTINGALE, P.G. AND NEUBERGER, J. (2005): 'Self-Organizing Maps Can Determine Outcome and Match Recipients and Donors at Orthotopic Liver Transplantation', *Transplantation*, **79**, pp. 213-218

[4] LISBOA, P.J.G. (2002): 'A review of evidence of health benefit from artificial neural networks in medical intervention', *Neural Network,* **15**, pp. 11-39

[5] HAYDON, G.H., JALAN, R., ALA-KORPELA, M., HILTUNEN, Y., HANLEY, J., JARVIS, L.M. AND LUDLUM, C.A. (1998): 'Prediction of cirrhosis in patients with chronic hepatitis C infection by artificial neural network analysis of virus and clinical factors', *J. Viral Hepatitis*, **5**, pp. 255-264

[6] KAARTINEN J., HILTUNEN, Y., KOVANEN, P.T. AND ALA-KORPELA, M. (1998): 'Classification of Human Blood Plasma Lipid Abnormalities by ^{1}H Magnetic Resonance Spectroscopy and Self-Organizing Maps', *NMR Biomed*. **11**, pp. 168-176

[7] PULKKINEN, J., LAPPALAINEN, M., HÄKKINEN, A.-M., LUNDBOM, N., KAUPPINEN, R.A. AND HILTUNEN Y. (2003): 'Quantification of Human Brain Metabolites from *In Vivo* ^{1}H NMR Magnitude Spectra Using Self-Organising Maps', *Lecture Notes in Computer Science*, **2690**, pp. 522-529

[8] KOHONEN, T. (2001): 'Self-organizing Maps', (Springer-Verlag, Berlin Heidelberg New York)

[9] MACQUEEN, J. (1967): 'Some methods for classification and analysis of multivariate observations', Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I: Statistics, 1967, pp. 281-297

ISSN: 1727-1983