Predicting Early Transplant Failure: Neural Network Versus Logistic Regression Models

Vicente Ibáñez[1,*], Eugenia Pareja[1], Antonio J. Serrano [2], Juan José Vila[1], Santiago Pérez[3], José D. Martín[2], Fernando Sanjuán[1], Rafael López[1] and Jose Mir[1]

[1]*Hepatic Surgery and Liver Transplant Unit. La Fe University Hospital (Valencia);* [2]*Digital Signal Processing Group. Department of Electronic Engineering. University of Valencia and* [3]*Valencian School for Health Studies (EVES), Spain*

Abstract: Cox's proportional hazard model or logistic regression model has been the classical mathematical approach to predict transplant results, but artificial neural networks may offer better results. In order to compare both methods, a logistic regression and a neural network model were generated to predict early transplant failure assessed at 90 days.

Methods: Medical charts from 701 liver transplant patients were used as generation cohort, collecting variables from donor, recipient and operative data. The discrimination capacity of the models was measured through the area under their ROC curves. Models were validated by applying them to a second cohort of 170 patients (validation cohort), although afterwards it was enlarged to 246 patients in order to increase statistical power.

Results: For the generation sample, ROC curves were 75% for logistic regression and 96% for neural network (χ^2 = 44,60. p<0,00001). Applied to the whole validation sample these values dropped to 68.7 % for logistic regression and 69.9 % for neural network (χ^2 = 0.026. p: 0,87). However, when models where applied to the validation cohort in cumulative groups of 50 patients two aspects became evident: 1) predictions worsened for patients who were more distant in time from the generation cohort; 2) for the first hundred patients in validation cohort, neural network was clearly superior to logistic regression model (93 % vs 76 %; χ^2 = 10.52. p:0,001).

Conclusions: Our results suggest that, provided with the same information and for a limited period of time, neural networks may offer better diagnostic performances than with logistic regression models.

Key Words: Liver transplantation, statistical models, artificial neural network.

INTRODUCTION

The role of predictive models in liver transplants has evolved quite substantially in recent years, fuelled mainly by the increasing lack of donors. This situation was foreseen by Starzl back in 1989, when he predicted that the organ supply would increasingly influence the candidacy criteria and would limit the practice of this procedure [1]. The earliest models, initially developed to identify outcome-related variables such as primary failure or graft dysfunction [2-6], have evolved and are now used as tools for managing the allocation of grafts according to their likelihood of success [7-9].

Cox's proportional hazard model and the logistic regression model have become the most commonly used mathematical approaches to solving this problem. Neural networks (NN) are a less popular alternative. The name 'neural network' alludes to its similarity with the human brain: the method comprises a structure of basic and interconnected elements (artificial neurons) that have to be trained in order to yield a suitable answer. This method implies that the system is, in fact, able to learn, and it is preferred over other mathematical methods when solving certain problems, such as when it is difficult to find a set of defining systematic rules, or when the problem presents changeable conditions. Neural networks are also very powerful when there is a significant number of sample cases or when there is a large number of explicative variables [10], all of which are present in liver transplants. Another advantage of this methodology is that previous hypotheses are not required about the system from which the information is to be sourced, unlike generalised linear models which assume an exponential distribution function. One of the classical criticisms of neural networks is that compared to conventional statistical techniques, they have a lower explanatory power, resembling "black boxes" that are unable to explain how the decisions are taken. However, a variety of authors have demonstrated that it is also possible to extract knowledge from NN [11,12].

With regard to liver transplantation, this method was already applied when studying different range of aspects, as recurrence of hepatocellular carcinoma [13], early graft failure [14], or survival rates at three or twelve months [15]. In light of the theoretical advantages of neural networks over traditional methods, and as it is suggested that in the near future the method for allocating grafts shall be based on the likelihood of success [16], our working hypothesis is that using the same data, a model developed using NN will be more able to discriminate than other model developed by means of logistic regression.

*Address correspondence to this author at the Pediatric Surgery Unit, Hospital General de Castellón, Avda. Benicàssim, s/n, 12004 Castellón, Spain; Tel: +34 964 726 500; Fax: +34961973286;
E-mail: ibanez_vic@hotmail.com

1874-4184/09

MATERIAL AND METHODS

All data for generating predictive models were obtained from charts from the Liver Surgery and Transplant Unit of La Fe University Hospital in Valencia, Spain. Files pertaining to transplants carried out between August 1997 and August 15th 2004 were used. Multi-organ transplants, paediatric transplants (recipient under 14 years of age) and those carried out with special techniques (split liver, living donor) were excluded. Readily available and preferably objective variables were selected, with the condition they were previously related to the result by means of multivariant type studies (Table **1**). The number of variables employed was limited in order to reduce the risk of random associations, which rises with the number of variables included [17]. The MELD score was not considered as it was not included in the majority of patients in the generation sample, and we preferred not to estimate INR based on the available data. Even though a recent study show the influence of the MELD in the result of the transplant during the first year [18], its absence in both models minimises its impact on the study aim.

Table 1. Variables Selected for Generating the Models

Donor	Recipient	Operative
Age	Age	Cold ischemia
Gender	Gender	Warm ischemia
Cause of death	Transplant indication	Blood consumption
Days in intensive care	UNOS stage	
Need for vasopressors	Serum creatinine	
pH	Bilirubin	
Steatosis	Protrombin time	
Serum sodium	Child-Pugh score	

Transplant failure assessed at 90 days was the main outcome, considering failure being both death of the recipient from directly related causes as retransplantation due to primary failure.

Logistic Regression

The logistic regression model (LR) was developed at the Valencian School for Health Studies. Previous univariant analysis were used, including in the multivariant analysis those variables with a $p < 0.10$. The model was adjusted using the SPSS 11.0.1. statistical program according to classical procedures. The end variables of the final model were selected using the procedure of sequential input based on the likelihood ratio (forward selection by LR), and comparing the result independently by repeating the process with the procedure of sequential withdrawal (backward selection by LR) with input probabilities of 0.05 and output probabilities of 0.1. Finally, the most balanced model is selected formed only with the variables whose coefficients showed statistical significance, the level of which was set to the standard 5%.

Neural Network

The neural network model (NN) was trained at the Electronic Engineering Department of the University of Valencia. The NN was developed using a specific module belonging to the Digital Signal Processing Group of this Department. Programming was carried out on a Matlab environment (belonging to MathWorks).

The type of neural network used was the *multilayer perceptron* with two layers of neurons (Fig. **1**). The hyperbolic tangent function was used at the hidden and output nodes as activation function, since it improves training performance in problems related to sample classification.

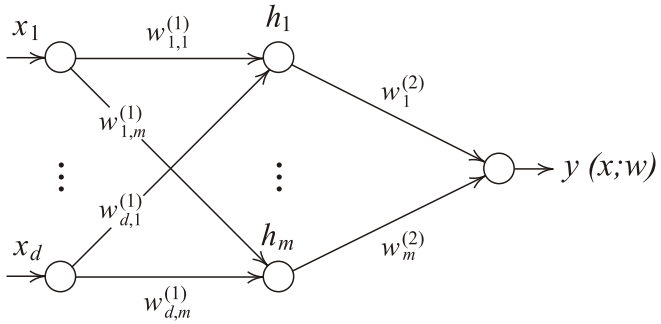


Fig. (1). Schematic representation of the multilayer perceptron structure.

Each training iteration was carried out by applying the *backpropagation algorithm*. As the algorithm's name implies, the errors (and therefore the learning) propagate backwards from the output nodes to the input nodes. We varied the number of hidden neurons from 2 to 10 (<10 to avoid overfitting). Data were divided at random in two groups, one with 25% of sample size to constantly evaluate the level of overfitting or memorization of data used in learning. Models were selected by evaluating the sum of the sensitivity and specificity factors to obtain well-balanced models.

The discrimination capacity of the models was measured through the area under their ROC curves (AUC). Comparisons between obtained AUCs was carried out in accordance with Hanley and McNeil's [19] method, considering both curves correlated, using the EpiDat 3.1. program.

Models were validated by applying them to a second set of patients transplanted between August 16th 2004-April 2006. For selecting the cutoff point of the ROC curves to be used in practice it was assumed that the cost of a false negative (the model indicates success but the patient dies) was worse than a false positive.

RESULTS

Generation Sample

During the period of the study 729 patients were operated on, with 701 fulfilling inclusion criteria. 101 transplants failed in the first 90 days, which accounts for an early failure prevalence of 14.4% (IC 95% 11.73-17.07). Sample characteristics were similar to the rest of the Spanish enlisted population, with a predominance of transplants carried out because of post viral cirrhosis and alcoholic etiology, accounting for 77% (Table **3**). Survival was similar to that of a wider series, with survival rates of 85.6% at 90 days, increasing to 88.35% if retransplants are excluded. These values are simi-

Table 2. Characteristics of Samples for Continuous Variables. Student's T-Test or Mann-Withney U-Test was Used, Depending on Normality Test

	GROUP				p
	Generation (n=701)		Validation (n=246)		
	Mean	S.D.	Mean	S.D.	
Bilirubin (mg/dl)	5,17	8,05	6,81	10,5	**0.058**
Creatinin (mg/dl)	1,07	0,60	0,95	0,72	**0.0001**
ICU stay (days)	2,64	2,86	3,23	4,53	**n.s.**
Donor age (years)	46,86	18,62	50,19	19,23	**0.0001**
Recipient age (years)	52,74	10,73	53,18	9,63	**n.s.**
Warm ischemia (minutes)	39,49	17,23	43,75	14,73	**0.003**
Cold ischemia (minutes)	363,75	168,51	335,84	164,19	**0.028**
Donor pH	7,41	0,09	7,40	0,09	**0.013**
Child-Pugh score	8,94	2,18	9,46	2,36	**0.006**
Protrombin time (%)	62,29	21,19	58,96	22,09	**0.018**
Blood consumption (units)	3,77	2,99	2,77	2,31	**0.001**
Donor sodium (mEq/l)	147,98	9,97	149,42	10,56	**n.s.**

lar to those obtained from the European Liver Transplant Registry (ELTR) for more than 34,000 cases [9].

Variables finally included in the LR model are shown in Table **4**. This model showed an AUC of 75%. With regard to the NN, its performance was clearly superior, with an AUC of 96% (Fig. **2**). Difference between AUCs was statistically significant ($\chi 2 = 44{,}60$. $p<0{,}00001$).

Validation Sample

170 charts were used for validation, from a whole sample of 187 transplants performed in the second period of the study. Twelve transplants failed during the first 90 days, which means a prevalence of 7.05% (IC 95%: 2.91-11.20), significantly lower than the one founded in the generation sample (z: 2.43, p: 0.015).

The discrimination capacity of the LR model was 78%, whereas for the NN model it reached 81% (Fig. **2**). In this case the difference between AUCs did not reach statistical significance ($\chi^2 = 0.09$. p: 0.75). Using the method proposed by Hanley and McNeil [20] to compare curves derived from same cases, for an α error of 5% and a statistical power of 80%, 21 cases would be needed for each result (dead/alive), whereas in the sample studied we only had seven cases in the "failure" group, as a consequence of charts with missing values, and only cases with complete data can generate a prediction from the models. In an attempt to obtain more failed cases, the validation sample was expanded to 246 cases out of the 266 transplants performed until February 2007. Results were very similar for both groups: LR model 68.7%, NN model 69.9%; ($\chi^2 = 0.026$, p: 0.87) (Fig. **2**). The prevalence of early failure in this second validation cohort (22 cases, 8.9%) was also significantly lower than the generation sample (z: 2.083, p: 0.037).

Given the fact that the results of the generation sample were clearly favourable to the NN, one possible explanation is that the diagnostic performance in the more recent cases was worse than in the older cases – which were thus closer in time to the generation sample – so that the results of both models were compared according to the size of the validation sample. Table **5** shows the results, revealing how the inclusion of new cases penalises the performance of the NN, which initially performed much better than the LR; however, with more than 100 cases there are no statistically significant differences between both models.

DISCUSSION

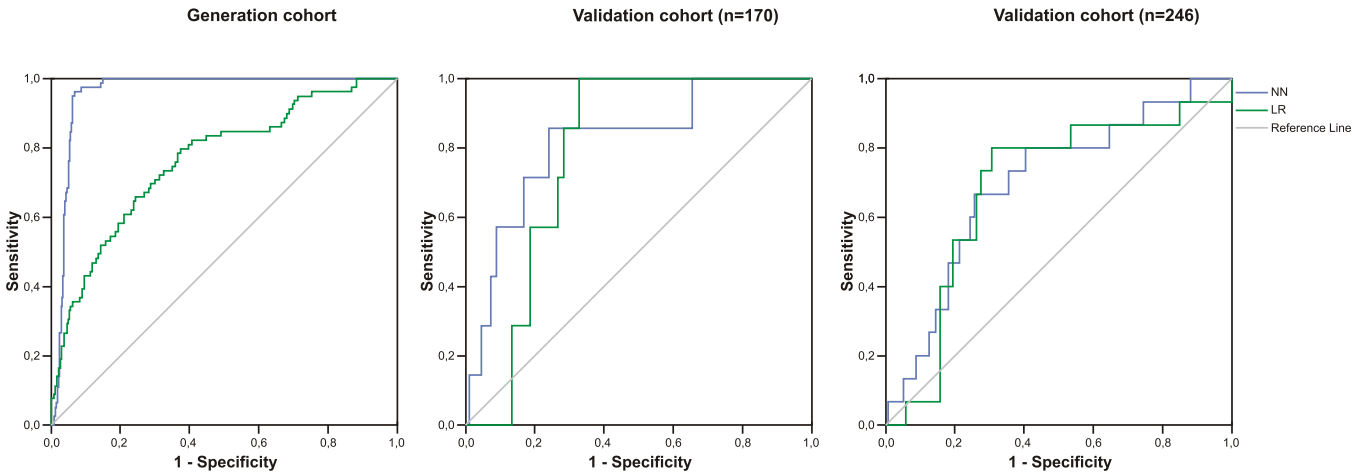
In spite of the different options developed to increase the number of available grafts, there is still a gaping chasm worldwide between the organ supply and demand which results in the need to rationally manage a scarce and expensive resource. Each transplant failure means the loss of a valuable resource and may mean the death of another patient on the waiting list. Under these conditions, rationally using organs means offering them first to patients on the waiting list with a lower chance of survival; however, the likelihood of success -or the likelihood of failure- also needs to be taken into account. Still, the idea of assigning organs only to patients with the greatest likelihood of success is rather questionable from an ethical point of view, since transplant results should be interpreted from an intent-to-treat basis given these patients' dismal prognosis. However, in the event that

Table 3. Characteristics of Samples for Categorical Variables. Comparisons Through Chi-Square Test

	GROUP				p
	Generation (n=701)		Validation (n=246)		
	n	%	n	%	
Donor cause of death					
trauma	239	34,6	69	28,4	**n.s.**
stroke	415	60,1	163	67,1	
anoxia	37	5,4	11	4,5	
Transplant indication					
postnecrotic	541	77,4	184	74,8	**n.s.**
cholestatic	29	4,1	7	2,8	
malignancy	16	2,3	4	1,6	
acute failure	19	2,7	11	4,5	
miscellaneous	39	5,6	22	8,9	
early retransplant	28	4	8	3,3	
late retransplant	27	3,9	10	4,1	
UNOS					
home	612	87,4	208	86,3	**0.01**
hospital	59	8,4	13	5,4	
ICU	29	4,1	20	8,3	
Vasopressors					
No	463	66	110	44,7	**0.0001**
Yes	238	34	136	55,3	
Sex match					
Male-male	310	44,2	117	47,8	**n.s.**
Male-female	112	16,0	30	12,2	
Female-male	173	24,7	67	27,3	
Female-female	106	15,1	31	12,7	
Steatosis					
mild	696	99,3	237	96,3	**0.0001**
moderate	5	0,7	8	3,3	
severe	-	-	1	0,4	

there were an appropriate instrument, donor-recipient combinations that show a high likelihood of failure could be avoided.

The fist part of this problem has been resolved by applying the Model for Endstage Liver Disease (MELD), which was shown to be more reliable than the Child-Turcotte-Pugh Score as the instrument to prioritise organ allocation [20]. However, MELD has been shown to be somewhat inaccurate when applied to transplant results. Consequently, a new model taking into account both donor and recipient characteristics would be needed [18]. To date, no predictive model has been developed that is sufficiently reliable to be used in practice, although there have been multiple theoretical attempts. The models developed based on Cox's regression



	AUC	CI 95%	AUC	CI 95%	AUC	CI 95%
NN	0,96	0,94-0,97	0,81	0,65-0,97	0,69	0,56-0.83
LR	0,76	0,70-0,82	0,78	0,69-0,86	0,68	0,54-0,82
	X^2	p	X^2	p	X^2	p
Comparison between AUCs	44,6	0,00001	0,099	0,75	0,026	0,87

Fig. (2). Diagnostic performance of predictive models.

Table 4. Variables Included in the Logistic Regression Model

Variable	Odds Ratio	S. E.	p	95% C.I.
Child-Pugh score	1.288	0.112	0.004	1.086 - 1.527
Protrombin time	1.026	0.009	0.005	1.007 - 1.045
Blood consumption	1.209	0.048	0.000	1.117 - 1.308
Bilirubin	1.050	0.017	0.003	1.017 - 1.086
Donor Vasopressors	1.665	0.440	0.054	0.991 - 2.795
Sex combination (donor-recipient)				
Male-female	1.446	0.521	0.306	0.713 - 2.932
Female-male	0.716	0.249	0.337	0.362 - 1.415
Female-female	2.098	0.735	0.035	1.055 - 4.172
Cold ischemia time	1.001	0.0007	0.023	1.0002 - 1.003

[7,8,18] have the problem of offering a relative as opposed to an absolute risk of failure, which entails the need to previously set up a threshold beyond which transplants should be rejected [16]. One alternative to this methodology is logistic regression [9], which, unlike Cox's regression, provides the likelihood of death for each specific case, thus enabling it to be used as a binary diagnostic test that classifies patients (dead/alive or success/failure) once established a cut-off point to determine class membership. Neural network models fall within this latter group, and they have been successfully applied in a variety of fields where the interactions between co variables are significant, such as in classification, modelling and signal processing problems [21]. They have also been used to predict temporary series in diverse fields such as economics [22], engineering [23], telecommunications [24] and medicine, where, among other areas, they have been applied in liver transplantation to predict different outcomes such as primary failure, death, tacrolimus levels or cancer recurrence [14,15,25-27]. More recently, a NN model applied on waiting list mortality [28] has showed better results than the MELD, generated by means of a Cox's regression.

In our study, NN showed a higher ability to discriminate than LR in the generation sample, but no in the validation

Table 5. AUC comparisons between Models Depending on Validation Sample Size

Cases in Validation Sample	AUC (CI 95%)		Comparison between AUC	
	NN	LR	X^2	p
50	0.957 (0.88-1)	0.774 (0.62-0.92)	6.599	0.010
100	0.938 (0.87-0.99)	0.768 (0.66-0.87)	10.522	0.001
150	0.817 (0.66-0.97)	0.798 (0.71-0.88)	0.037	n.s.
200	0.731 (0.56-0.89)	0.693 (0.51-0.86)	0.097	n.s.
246	0.699 (0.56-0.83)	0.687 (0.54-0.82)	0.026	n.s.

sample, where both models' results were quite similar. However, when results were analyzed depending on sample size, NN offered figures over 90%, an unusual finding in this field. Both models' gradual loss in the power to discriminate as the size of the validation cohort rose is compatible with their being applied to subjects that are increasingly distant in time from those that generated it, with up to ten years of difference. This loss in precision is more pronounced in the NN model due to its better results in the first 100 cases, dropping later to figures similar to those using LR. Another interesting point is the differing degrees of early failure prevalence between generation and validation samples, a finding for which we have no clear explanation. Comparison between groups (Table **2**) reveals that subjects in the validation sample had a lower cold ischemia time, blood consumption and creatinin, yet they also showed higher bilirubin levels, donor age and Child-Pugh scores. Early failure prevalence trends do not seem to fall either. In any case, we believe that these differences do illustrate real practice.

A variety of factors will have to be clarified before a useful model for clinical practice can be put forth, such as the number of variables to be included, and the mathematical method most appropriate to generate it. With regard to this point, our results, if confirmed in an external database, suggest that just like in other fields, neural networks may offer better results than the classical methods. We have used a very limited number of variables compared to other authors [15], but in practice the existence of computerised databases would enable a higher number of variables to be controlled and might make it possible to eliminate those that, such as ischemia times and operative blood consumption, cannot be known prior to surgery. Another factor to take into account is the way that these models might be used in assignment systems. Unlike the MELD, a universal model that is applicable to any patient, the use of a model for managing the assignment of grafts will depend on different circumstances in each clinical scenario, such as the number of donations available in each country, the early failure prevalence for each centre and the patient's clinical status. The result is that its utility will be variable and will depend on each specific case. For example, considering the utility of liver transplantation as a function depending on the expected outcome and the emergency, as proposed by Burton [29], a predictive model that offers 80% sensitivity and 88% specificity - figures offered by the NN model for the first 100 cases in the validation sample – in a hospital with a 9% prevalence of early failure, would present the utilisation strategy shown in Fig. (**3**). It shows how the predictive model would be the best option for patients with 90 days survival on the waiting list between 58% to 42%, which is equivalent to a MELD score between 14 to 19, respectively. This strategy would be different for a model with better sensitivity and specificity figures (the utility would rise), or if the centre had a 5% early failure rate (in this case, it would drop). Nor will the predictive model be either fixed or immobile, rather the ideal situation is for all the information available to be used for new predictions with the goal of gathering temporal trends that are difficult to quantify in concrete variables, such as overall improvements in patient management. In our study, a total of 947 records were managed (701 in the generation cohort and 246 in the validation cohort), but the prediction made for the last patient in the validation cohort took into account only the 701 transplants performed in the generation cohort, while in reality information on over 800 transplants was available. This would explain why the performance of both models was higher with the patients that were closer to the generation cohort and then gradually worsened.

Until today, when new technologies can solve the problem of the scarcity of organs, the use of predictive models may be an invaluable aid for rationally managing grafts. Nevertheless, proper use of these models, just like any other diagnostic test, will demand accurate evaluations of the pretest probabilities and the patient's clinical status, as stated by Pauker and Kassirer [30] more than 20 years ago.

ACKNOWLEDGEMENTS

Supported by grants from the Valencian School for Health Studies and from the Carlos III Health Institute (FIS grant 04/0453).

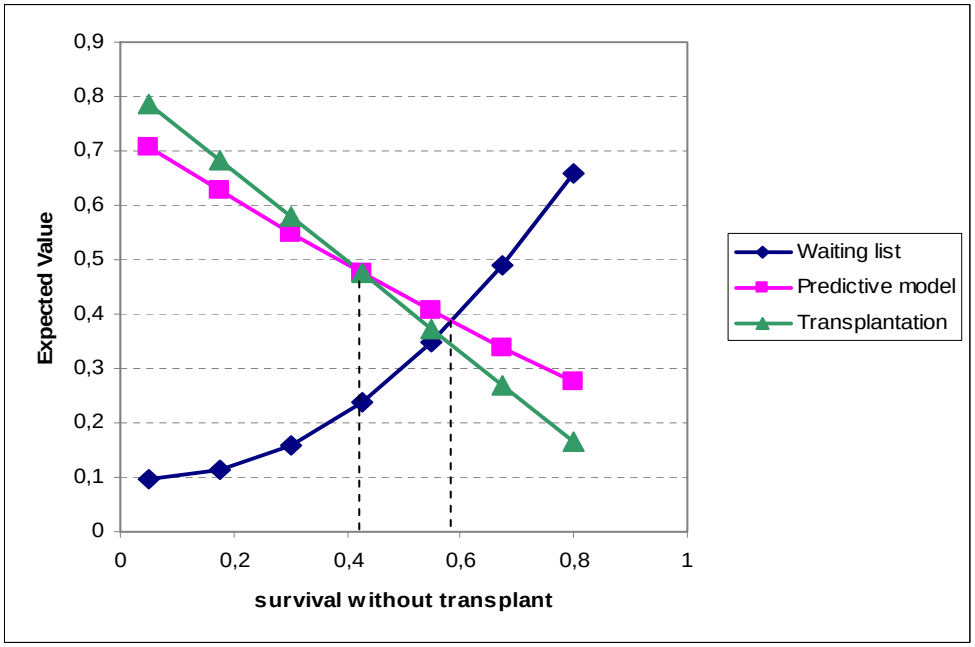


Fig. (3). Utility strategy for a theoretical predictive model with a 80% sensitivity and 88% specificity, applied in a centre with an early failure prevalence of 9%. For patients with a MELD between 14 and 19, the predictive model would be preferred.

ABBREVIATIONS

AUC	=	Area under curve
ELTR	=	European liver transplant registry
MELD	=	Model for end stage liver disease
LR	=	Logistic regression
NN	=	Neural network
ROC	=	Receiver operator characteristics
CI 95%	=	Confidence Interval at 95%
χ^2	=	Chi- square

REFERENCES

[1] Starzl TE, Demetris AJ, Van Thiel D. Liver transplantation (first of two parts). N Engl J Med 1989; 321: 1014.

[2] Baliga P, Merion RM, Turcotte JG, *et al.* Preoperative risk factor assessment in liver transplantation. Surgery 1992; 112: 704.

[3] Ploeg RJ, D'Alessandro AM, Knechtle SJ, *et al.* Risk factors for primary dysfunction after liver transplantation-A multivariate analysis. Transplantation 1993; 55: 807.

[4] Marino IR, Morelli F, Doria C, *et al.* Preoperative assessment of risk in liver transplantation: A multivariate analysis in 2376 cases of the UW era. Transplant Proc 1997; 29: 454.

[5] Hernández D, Jiménez C, Loinaz C, *et al.* Risk factors of graft loss in orthotopic liver transplantation. Transplant Proc 1998; 30: 3241.

[6] González E, Rimola A, Navasa M, *et al.* Liver transplantation in patients with non-biliary cirrosis: prognostic value of preoperative factors. J Hepatol 1998; 28: 320.

[7] Adam R, Cailliez V, Majno P, *et al.* Normalised intrinsic mortality risk in liver transplantation: European Liver Transplant Registry study. Lancet 2000; 356: 621.

[8] Ghobrial RM, Gornbein J, Steadman R, *et al*. Pretransplant model to predict postransplant survival in liver transplant patients. Ann Surg 2002; 236: 315.

[9] Burroughs AK, Sabin CA, Rolles K, *et al*. 3-month and 12-month mortality after first liver transplant in adults in Europe: predictive models for outcome. Lancet 2006; 367: 225.

[10] Martín B, Sanz A. Redes neuronales y sistemas borrosos. Madrid: Ra-Ma; 1997.

[11] Andrews R, Diederich J, Tickle AB. A survey and critique of techniques for extracting rules from trained artificial neural networks. Knowl-Based Syst 1995; 8: 373.

[12] Tickle AB, Andrews R, Golea M, Diederich J. The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. IEEE Trans Neural Networks 1998; 9: 1057.

[13] Rodriguez-Luna H, Vargas HE, Byrne T, Rakela J. Artificial neural network and tissue genotyping of hepatocelllar carcinoma in liver-transplant recipients: Prediction of recurrence. Transplantation 2005; 79: 1737.

[14] Doyle HR, Dvorchik I, Mitchell S, *et al.* Predicting outcomes after liver transplantation. A connectionist approach. Ann Surg 1994; 219: 408.

[15] Haydon GH, Hiltunen Y, Lucey MR, *et al.* Self-organizing maps can determine outcome and match recipients and donors at orthotopic liver transplantation. Transplantation 2005; 79: 213.

[16] Lucey MR. How will patients be selected for transplantation in the future?. Liver Transpl 2004; 10: S90.

[17] Chung SW, Kirkpatrick AW, Kim HLN, Scudamore CH, Yoshida EM. Correlation between physiological assessment and outcome after liver transplantation. Am J Surg 2000; 179: 396.

[18] Habib S, Berk B, Chang CCH, *et al.* MELD and prediction of post-liver transplantation survival. Liver Transplant 2006; 12: 440.

[19] Hanley JA, McNeil BJ. A Method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839.

[20] Wiesner RH, McDiarmid SV, Kamath PS, *et al.* MELD and PELD: Application of survival models to liver allocation. Liver Transplant 2001; 7: 567.

[21] Kornilov AR. Intelligent technologies: new opportunities for modern industry. Inform Technol 1997; 3: 1.

[22] Trigueiros D, Taffler R. Neural networks and empirical research in accounting. Account Bus Res 1996; 26: 347.

[23] Denby B. Neural networks in high energy physics: a ten years perspective. Comput Phys Commun 1999; 119: 219.

[24] Christiansen A, Herschtal A, Herzberg M, Kowalczyk A, Szymanski J. Neural Networks for Resource Allocation in Telecommunication Networks. Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications 3, IWANNT'97, Lawrence Erlbaum Associates, Publishers, Hillside, N.J., 1997; pp. 265-273.

[25] Dvorchik I, Subotin M, Marsh W, McMichael J, Fung JJ. Performance of multi-layer feedforward neural networks to predict liver transplantation outcome. Methods Inf Med 1996; 35: 12.

[26] Hughes VF, Melvin DG, Niranjan M, Alexander GA, Trull AK. Clinical validation of an artificial neural network trained to identify acute allograft rejection in liver transplant recipients. Liver Transplant 2001; 7: 496.

[27] Parmanto B, Doyle HR. Recurrent neural networks for predicting outcomes after liver transplantation: representing temporal sequence of clinical observations. Methods Inf Med 2001; 40: 386.

[28] Cucchetti A, Vivarelli M, Heaton ND, *et al*. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. Gut 2007; 56: 253.

[29] Burton Jr JR, Sonnenberg A, Rosen HR. Retrasplantation for recurrent hepatitis C in the MELD era: maximizing utility. Liver Transplant 2004; 10: S59.

[30] Kassirer JP, Pauker SG. Should diagnostic testing be regulated? N Engl J Med 1978; 229: 947.

Received: January 12, 2009 Revised: February 26, 2009 Accepted: March 26, 2009