ELSEVIER

Application of an Artificial Neural Network Model to Predict Delayed Decrease of Serum Creatinine in Pediatric Patients After Kidney Transplantation

G. Santori, I. Fontana, and U. Valente

ABSTRACT

Artificial neural network, a computer-based technology that uses nonlinear statistics to recognize the relationship between input variables and an output variable, has been previously applied to outcome prediction in adult kidney recipients. In this study, we evaluated the effectiveness of a neural network model to predict a delayed decrease of serum creatinine in pediatric kidney recipients. The neural network was constructed with a training set of pediatric kidney recipients ($n = 107$) by using 20 input variables and assuming for the output variable, the time after 3 days to reach a serum creatinine level 50% below that before kidney transplantation. In the final model, the following input variables showing higher predictive values were retained: serum creatinine on day 1 post transplant, urine volume in the first 24 hours, diagnostic category, pretransplant dialysis mode, patient sex, donor sex, body weight on day 1 posttransplant, and patient age. The model was validated in a second set of patients ($n = 41$) by blinding the network for the output variable. The overall accuracies of the neural network for the training set, the validation set, and the whole patient cohort were 89.1%, 76.92%, and 87.14%, respectively. A comparative logistic regression analysis revealed only serum creatinine on day 1 posttransplant to be an independent predictor for the output variable (overall accuracy: 79.05%). The neural network showed sensitivity and specificity for the whole patient cohort to be 0.875 and 0.87, respectively, whereas using logistic regression sensitivity and specificity yields 0.37 and 0.94, respectively. This study proposes a neural network model that seemed to predict a delayed decrease in serum creatinine among pediatric kidney recipients. The availability of the source code may allow development of stand-alone neural networks to validate our model in prospective studies.

DELAYED RECOVERY of graft function is a common complication of kidney transplantation that may affect both short- and long-term graft outcomes.[1–4] Although multivariate analyses may identify several independent predictors for early graft function, the attempt to make a prediction at the level of each kidney recipient represents a more difficult target. With this aim, application of artificial neural network (ANN) technology seemed to be an attractive strategy, considering that this technique uses pattern recognition to discover relationships between input and outcome variables.[5,6] Once constructed by entering historical data, a trained ANN model may be applied to patients with unknown outcomes. The ANN approach has been successfully used as a diagnostic and predictive tool in various clinical settings.[7] For example ANN technology was previously utilized to predict early graft function after grafting of adult kidney recipients.[8,9]

In the present work, a specific ANN model was evaluated in pediatric patients who underwent grafting from cadaveric donors, assuming delayed decrease of serum creatinine (DDSCr) as the output variable. The final ANN model was trained, validated, and then compared

From the Department of Transplantation, San Martino University Hospital, L.go R. Benzi 10, 16132 Genoa, Italy.

This study was supported by grants from Italian Ministry of University and Scientific Research.

Address reprint requests to Gregorio Santori, MD, PhD, Department of Transplantation, San Martino University Hospital, L.go R. Benzi 10, 16132 Genoa, Italy. E-mail: gregorio.santori@gmail.com

0041-1345/07/$–see front matter
doi:10.1016/j.transproceed.2007.05.026

for overall accuracy with a conventional logistic regression (LR) model.

PATIENTS AND METHODS

Data Collection

We retrospectively collected clinical data on pediatric patients (n = 257) with end-stage renal disease who underwent transplantation from cadaveric donors over a 171-month period. Pediatric kidney recipients on peritoneal dialysis (PD) or hemodialysis (HD) for at least 3 months[10] were included in the study. Patient exclusion criteria were: no switch in pretransplant dialysis mode, nonfunctional grafts, obstructive nephropathy, biopsy-demonstrated hyperacute rejection, posttransplant acute renal failure, multiple-organ transplant, kidney retransplantation, or an incomplete pre- or posttransplantation dataset.[10] Following these criteria, 148 patients were enrolled in the analysis.

Diagnostic Categories

In agreement with the Medical Policy Technology Assessment Committee,[11] patient diagnostic categories were classified as follows: irreversible chronic renal failure (n = 50), hereditary nephropathies (n = 26), irreversible acute renal failure (n = 18), congenital disorders (n = 17), metabolic disorders (n = 6), obstructive uropathy (n = 3), toxic nephropathies (n = 2), tumors requiring nephrectomy (n = 1), other indications with a documented chronic renal failure of at least 6 to 8-week duration (n = 25).

Parameters

The following continuous patient-related variables were considered: age; height on the graft day; body weight (BW) on that day; body mass index (BMI) on that day, calculated as [weight (kg)/height (m)^2]; body surface area (BSA) on that day, calculated as previously described[12]; body weight on day 1 (BW1); body weight gain (BWG), calculated as the difference between BW on that day and on day 1; urine volume collected by an indwelling bladder catheter in the first 24 hours (UV1); serum creatinine (SCr) measured on the day 1 (SCr1); the number of days needed to reach a SCr level reduced by 50% [$T_{1/2(SCr)}$]. DDSCr was assumed for $T_{1/2(SCr)} > 3$ days, according to Van Biesen et al.[10] Patient-related variables treated as categorical variables were sex; pre-transplant dialysis mode (PDM); positivity for cytomegalovirus (CMV), Epstein-Barr virus (EBV), hepatitis C virus (HCV); and prior blood transfusions. PDM was chosen by nephrology units that had originally treated the pediatric patients, in agreement with the dialysis modality decision guide previously described.[13] Other parameters considered were cadaveric donor age, donor sex, cold ischemia time (CIT), and warm ischemia time (WIT).

Artificial Neural Network

The neural network was constructed using commercially available software (Statistica Neural Networks, StatSoft Inc., Tulsa, OK, USA). ANN was constructed and trained by using a first set of pediatric patients (n = 107) selected from the original cohort of kidney recipients on the basis of the prompt availability of the full dataset. The training sample size was defined in agreement with previous ANN applications in the field of transplantation.[8,9] All variables were entered as continuous or nominal inputs; diagnostic categories were converted by the software to numerical values before analysis. DDSCr was assumed as the output variable. $T_{1/2(SCr)}$ was not included in the input variables, being the reference parameter to assess DDSCr. In the first step, all variables collected for the training sample were entered in the ANN. The built-in Intelligent Problem Solver tool tested several network typologies, retaining the best network and the optimal set of input variables. Only variables that reached a sensitivity >1.00 in the best network were retained in the final model. Sensitivity analysis was performed according to the deterioration in modeling performance that occurred when each variable was no longer available to the model.[14] The best neural network was a multilayered perceptron network that was trained with back propagation.[15] Validation of the ANN was performed by entering remaining patients (n = 41) in which a full dataset was obtained subsequently, blinding the neural network for DDSCr. Finally, the ANN model was tested in the whole patient cohort (training/validation), by blinding the neural network for DDSCr in validation patients. The source code of the ANN was obtained in both Visual Basic and C++ languages by using the built-in Code Generator tool.

Logistic Regression Model

A conventional LR model was constructed by entering the whole set of variables from all patients, assuming DDSCr as the dependent variable. The variables that were statistically significant on univariate analysis were entered in the multivariate LR model. Quasi-Newton was the main estimation method for both univariate and multivariate LR; when quasi-Newton failed to reach convergence for reasonable estimates, Hooke-Jeeves pattern moves method was used to generate estimates.[16]

Statistical Analysis

Accuracy [100 · (true positive + true negative)/(true positive + true negative + false positive + false negative)], misclassification rate (MCR; 1 − overall fraction correct), odds ratio (OR), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios for positive tests (LR+), likelihood ratios for negative tests (LR−), kappa, Forbes' normalized mutual information index (NMI)[17] and Youden's J (sensitivity + specificity − 1)[18] were calculated with their 95% confidence interval (CI) for both ANN and LR models. Receiver-operating characteristic (ROC) curves were generated for each neural network. The area under the ROC curve may range from 0 to 1, with 1 corresponding to perfect discrimination and 0.5 what is expected by chance alone.[19] The accuracy between ANN and LR was compared for equal sample sizes by McNemar's test. Continuous and categorical variables were compared by Mann-Whitney test and chi-square test with Yates' continuity correction, respectively. Continuous variables were reported as mean values ± standard deviations, medians, and 95% CI. Statistical significance was assumed for a P value < .05 with a two-tailed null hypothesis. Statistical analyses were performed using the software package STATISTICA 7.1—Statistica Neural Networks (StatSoft Inc, Tulsa, OK, USA), MedCalc 7.5 (MedCalc Software, Mariakerke, Belgium), and the calculators Bayesian Statistics II MultiCalc (by L. Leff, MedCalc3000) and 2-way contingency table analysis (by J. Pezzullo, in http://members.aol.com/johnp71/javastat.html).

RESULTS

The overall dataset including both continuous and categorical variables is summarized in Table 1. The incidence of

Table 1. Variables Collected to Construct the Neural Network Model

		Median	95% CI
Patient continuous variables			
Patient age (y)	14.6 ± 5.39	15	13.72–15.47
Height (cm)	139 ± 26	146	134–143
BW (kg)	39.1 ± 16	40.5	36.47–41.75
BSA (m^2)	1.21 ± 0.35	1.27	1.15–1.27
BMI (kg/m^2)	19.6 ± 5.7	18.5	18.7–20.6
BW1 (kg)	40.9 ± 16.36	42.2	38.19–43.58
BWG (g/24 h)	1834 ± 2470	1000	1432–2235
SCr1 (mg/dL)	6.16 ± 3.1	5.9	5.65–6.67
UV1 (mL/24 h)	3109 ± 2540	2730	2694–3525
$T_{1/2(SCr)}$ (d)*	3.63 ± 3.65	2	3.02–4.23
Other continuous variables			
Donor age (y)	14.3 ± 10.9	12	12.53–16.08
CIT (h)	15.23 ± 3.53	15	14.66–15.81
WIT (min)	46.52 ± 9.75	45.5	44.91–48.12
Patient categorical variables			
Sex (males)	77 (52.01%)		
Pretransplant dialysis mode			
PD	64 (43.24%)		
HD	84 (56.75%)		
CMV (+)	144 (97.29%)		
EBV (+)	145 (97.97%)		
HCV (+)	4 (2.7%)		
Blood transfusions (+)	12 (8.1%)		
Other categorical variables			
Donor sex (males)	102 (68.91%)		
DDSCr [$T_{1/2(SCr)} > 3$ d]	40 (27.02%)		

*Variable used to define DDSCr (output variable); CI, confidence interval; BW, body weight on the day of kidney transplantation; BSA, body surface area; BMI, body mass index; BW1, body weight on the day 1 after kidney transplantation; BWG, body weight gain; SCr1, serum creatinine measured on the day 1 after kidney transplantation; UV1, urine volume collected in the first 24 hours after kidney transplantation; $T_{1/2(SCr)}$, the number of days needed to reach a serum creatinine level 50% below that before kidney transplantation; CIT, cold ischemia time; WIT, warm ischemia time; PD, peritoneal dialysis; HD, hemodialysis; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HCV, hepatitis C virus; DDSCr, delayed decrease of serum creatinine.

DDSCr was 27.02%. Comparison of characteristics between training and internal validation group showed statistical significance for BWG ($P < .001$), CIT ($P = .036$), HD ($P = .035$), and HCV+ ($P = .009$; Table 2). By comparison of training and internal validation sets for diagnostic categories, statistical significance occurred only for toxic nephropathies ($P < .0001$; Table 3).

To construct the ANN model, variables showed in Table 1 (including diagnostic category) were preliminarily entered for the training group ($n = 107$), while DDSCr was assumed as a dichotomous ("yes/no") outcome variable. $T_{1/2(SCr)}$ was excluded by the input variables, being the reference parameter to assess DDSCr.

The ANN underwent a first test/training process, in which the Intelligent Problem Solver tool tested different network typologies with the aim to retain the best network and the optimal set of input variables. Following this strategy, the best ANN was a multilayered perceptron network that showed an overall accuracy of 82.47% (MCR: 0.1753; OR: 19.33; sensitivity: 0.66; specificity: 0.9; PPV: 78.57%; NPV: 84.05%; LR+: 7.11 LR−: 0.36; kappa: 0.59; Forbes' NMI index: 0.27; Youden's J: 0.57; ROC-area: 0.83), with eight input variables that had major roles to predict the occurrence of DDSCr (Table 4). Sensitivity analysis was performed according to weakening in modeling performance that occurred when each variable was no longer available to the model. Only variables that reached a sensitivity value >1.00 in the best network were retained in the final model. In descending order of importance these variables were: SCr1, UV1, diagnostic category, PDM, patient sex, donor sex, BW1, and patient age. When the ANN model was trained after retaining only the input variables with sensitivity >1.00, the overall accuracy and sensitivity reached 89.1% and 0.86, respectively, with an improved whole predictive performance (Table 5). Trained ANN was a multilayer perceptron that showed a 8:14-10-1:1 structural design (training error: 0.135), with one hidden layer having 10 nodes. When the remaining patients that were not selected for training ($n = 41$) were entered for the input variables with higher sensitivity in the training group by blinding the neural network for DDSCr, the ANN model showed an overall accuracy of 76.92% (Table 5). When the same variables were entered for all pediatric patients (training and validation set) by blinding the neural network for DDSCr in the validation set, the predictive accuracy of ANN was 87.14% (Table 5). The Visual Basic and C++ source code of both trained and validated ANN was generated by a built-in Code Generator software tool.

A conventional univariate LR was performed by entering each variable collected in the whole pediatric patient cohort, with DDSCr as the dependent variable (Table 6). BW, BSA, BMI, BW1, SCr1, and WIT, which yielded significant values on univariate analysis, were entered as independent variables in a multivariate LR model. SCr1 was the unique variable to reach significance in the multivariate LR analysis [$\chi^2 = 31.53$; OR: 0.75 (0.64 to 0.88); $P < .0001$)]. The LR showed an overall accuracy of 79.05%, while sensitivity and specificity were 0.375 and 0.944, respectively (Table 5). In the whole pediatric patient cohort, the predictive accuracy of the LR was significantly lower than ANN model (79.5% vs 87.14%, $P = .043$).

DISCUSSION

A delay in functional recovery of transplanted kidneys negatively affects both short- and long-term graft outcomes.[1,2,4] Although conventional multivariate models identify several risk factors for the occurrence of delayed graft function (DGF),[4,10,20] they have failed to predict the outcome at the level of each kidney recipient, even when relevant risk factors were known.[9] In recent years, neural networks have emerged as an extension of conventional techniques in statistical pattern recognition, with successful applications in several medical areas.[5] The ANN approach

Table 2. Comparison of Continuous and Categorical Variables in Pediatric Kidney Recipients Grouped for Training and Internal Validation Set

	Training Set (n = 107)	Internal Validation Set (n = 41)		P Value‡
Patient continuous variables				
Patient age (y)	14.08 ± 4.9	15.95 ± 6.37		.08
Height (cm)	138 ± 26	141 ± 26		.601
BW (kg)	39.1 ± 16.4	39.1 ± 15.3		.8
BSA (m^2)	1.21 ± 0.36	1.22 ± 0.35		.727
BMI (kg/m^2)	20.03 ± 6.3	18.56 ± 3.12		.494
BW1 (kg)	40.4 ± 16.8	42 ± 15.36		.467
BWG (g/24 h)	1411 ± 2026	2936 ± 3135		<.001
SCr1 (mg/dL)	6.46 ± 3.01	5.35 ± 3.26		.115
UV1 (mL/24 h)	3238 ± 2768	2779 ± 1819		.547
$T_{1/2(SCr)}$ (d)*	3.6 ± 3.87	3.71 ± 3.02		.054
Other continuous variables				
Donor age (y)	13.4 ± 9.8	16.6 ± 13		.175
CIT (h)	15.59 ± 3.78	14.29 ± 2.58		.036
WIT (min)	47.22 ± 9.73	44.75 ± 9.72		.208
	Training Set (n = 107)	Internal Validation Set (n = 41)	χ^2	P Value§
Patient categorical variables				
Sex (males)	57 (53.27%)†	20 (48.78%)†	0.041	.839
Pretransplant dialysis mode				
PD	54 (50.47%)	10 (24.39%)	1.36	.242
HD	53 (49.53%)	31 (75.61%)	4.43	.035
CMV (+)	106 (99.07%)	38 (92.68%)	2.74	.097
EBV (+)	106 (99.07%)	39 (95.12%)	0.70	.400
HCV (+)	2 (1.87%)	2 (4.88%)	6.77	.009
Blood transfusions (+)	6 (5.61%)	6 (14.63%)	0.205	.65
Other categorical variables				
Donor sex (males)	75 (70.09%)	27 (65.85%)	0.057	.811
DDSCr [$T_{1/2(SCr)}$ >3 days]	30 (28.04%)	10 (24.39%)	0.027	.869

BW, body weight on the day of kidney transplantation; BSA, body surface area; BMI, body mass index; BW1, body weight on the day 1 after kidney transplantation; BWG, body weight gain; SCr1, serum creatinine measured on the day 1 after kidney transplantation; UV1, urine volume collected in the first 24 hours after kidney transplantation; $T_{1/2(SCr)}$, the number of days needed to reach a serum creatinine level 50% below that before kidney transplantation; CIT, cold ischemia time; WIT, warm ischemia time; PD, peritoneal dialysis; HD, hemodialysis; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HCV, hepatitis C virus; DDSCr, delayed decrease of serum creatinine.

*Variable used to define DDSCr (output variable); †within group %; ‡Mann-Whitney test; §chi-square test with Yates' continuity correction.

is a computer-based technique that uses nonlinear statistical analysis to discover the relationship between input and outcome variables.[5,6] Neural networks offer the benefit of generating models that are testable and predictive at a single-patient level, even if they cannot determine risk ratios for a whole patient cohort.[9] ANN technology has been applied to adult kidney recipients for both outcome prediction[9,21] and stratification of cardiac risk.[22] However, the use of ANN technology on pediatric kidney recipients has been negligible to date.

Table 3. Comparison of Diagnostic Categories in Pediatric Kidney Recipients Grouped for Training and Internal Validation Set

	Training Set (n = 107)	Internal Validation Set (n = 41)	χ^2	P Value†
ICRF	40 (37.38%)*	10 (24.39%)*	0.16	.689
HN	19 (17.76%)	7 (17.07%)	0.291	.589
IARF	12 (11.21%)	6 (14.63%)	0.305	.581
CD	13 (12.15%)	4 (9.76%)	0.582	.445
MD	2 (1.87%)	4 (9.76%)	1.76	.183
OU	—	3 (7.32%)	—	—
TN	1 (0.93%)	1 (2.44%)	29.27	<.0001
TRN	—	1 (2.44%)	—	—
OI	20 (18.69%)	5 (12.20%)	0.101	.751

ICRF, irreversible chronic renal failure; HN, hereditary nephropathies; IARF, irreversible acute renal failure; CD, congenital disorders; MD, metabolic disorders; OU, obstructive uropathy; TN, toxic nephropathies; TRN, tumors requiring nephrectomy; OI, other indications with a documented chronic renal failure of at least 6 to 8-weeks duration.

*Within group %; †chi-square test with Yates' continuity correction.

In this study, we developed a neural network with the aim to obtain a useful predictive tool for pediatric kidney recipients. Pediatric patients were enrolled by adopting a wide set of exclusion criteria to avoid potential bias caused by the cooccurrence of various confounding factors in the same patients, for example switch from one pretransplant dialysis modality to the other, multiple-organ transplant, acute renal failure, and/or retransplantation. DDSCr was assumed to be the output variable, considering that the SCr trend after transplantation is a crucial point in the early phase,[10,23] and that in our series long-term patient survival (≥3 years) was more than 92%. Different from previous applications of ANN technology, in which DGF was assumed as the output variable by adopting the broad defini-

Table 4. Sensitivity Analysis for Input Variables in the Best Test/Training Neural Network Model

Variable	Sensitivity
SCr1	1.298
UV1	1.244
Diagnostic category	1.086
Pretransplant dialysis mode	1.079
Patient sex	1.055
Donor sex	1.041
BW1	1.034
Patient age	1.014
BWG	1.000
CMV	1.000
HCV	1.000
Blood transfusions	1.000
BSA	0.997
BW	0.993
Height	0.973
Donor age	0.971
BMI	0.969
CIT	0.932
WIT	0.912

SCr1, serum creatinine measured on the day 1 after kidney transplantation; UV1, urine volume collected in the first 24 hours after kidney transplantation; BW1, body weight on the day 1 after kidney transplantation; BWG, body weight gain; CMV, cytomegalovirus; HCV, hepatitis C virus; BSA, body surface area; BW, body weight on the day of kidney transplantation; BMI, body mass index; CIT, cold ischemia time; WIT, warm ischemia time.

tion of dialysis need and/or increasing SCr level,[9] or by combining DGF with acute renal failure,[21] we have assumed DDSCr for $T_{1/2(SCr)} > 3$ days as a more stringent and measurable outcome, according to Van Biesen et al.[10]

The ANN was built with a training set of patients selected from the original cohort of pediatric kidney recipients. The training sample size included more than 100 patients, according to the magnitude of previous ANN application in this field.[8,9] Although our analysis was performed as a retrospective study, the whole dataset was not available at the same time. Therefore, the prompt availability of a full dataset was adopted as a selection criterion for training sample definition, instead of a more conventional random selection of patients.[9,24] This strategy has produced a training sample that showed differences after comparison with an internal validation set (Table 2), by making the latter group more similar to a theoretically external validation set. In previous ANN models where an internal validation set was selected at random from the original cohort of patients, no difference occurred between training and validation patients while impressive differences were observed with external patients.[24]

Although many variables were preliminarily entered into ANN as input variables (Table 1), cadaveric donor-related variables were limited to age and sex, considering that only optimal donors were selected for pediatric patients. HLA match was not included because this parameter is regulated by specific guidelines from Italian Ministry of Health and macroregional organ procurement organizations. By testing and training the ANN for the whole set of input variables ($n = 20$), the built-in Intelligent Problem Solver tool obtained the best combination of input variables with the highest predictive accuracy (Table 4). Following this strategy, previously described to predict mortality of cirrhotic patients,[24] eight variables were retained to construct the final ANN model. Although our trained ANN showed an high predictive performance (Table 5), a potential problem concerning neural networks is overtraining. An overtrained neural network may model the test group so fine that it becomes poor at predicting outcomes in new cases.[22,24] Our internal validation set, as described above, may be considered partly similar to an external validation set and produced a decrease in ANN overall accuracy when the network was masked for training patients, although we

Table 5. Predictive Performance of the Artificial Neural Network (ANN) Training Set, ANN Internal Validation Set, ANN Training/Internal Validation Set, and Logistic Regression Model for Delayed Decrease of Serum Creatinine (95% Confidence Interval)

Parameter	ANN Training Set ($n = 107$)	ANN Internal Validation Set* ($n = 41$)	ANN Training/Internal Validation Set* ($n = 148$)	Logistic Regression ($n = 148$)
Accuracy (%)	89.1 (81.8–93.4)	76.92 (63.5–84.1)	87.14 (81–90)	79.05 (73.6–83.1)
MCR	0.109 (0.066–0.182)	0.231 (0.159–0.365)	0.128 (0.091–0.19)	0.209 (0.169–0.264)
OR	59.42 (16.47–212)	12.57 (2.34–64.66)	46.84 (15.86–137.14)	10.2 (3.68–28.1)
Sensitivity	0.867 (0.744–0.939)	0.8 (0.538–0.939)	0.875 (0.768–0.941)	0.375 (0.274–0.449)
Specificity	0.9 (0.85–0.932)	0.759 (0.668–0.807)	0.87 (0.827–0.896)	0.944 (0.907–0.972)
PPV	0.788 (0.677–0.854)	0.533 (0.358–0.626)	0.729 (0.64–0.784)	0.714 (0.522–0.855)
NPV	0.941 (0.887–0.973)	0.917 (0.807–0.975)	0.946 (0.899–0.974)	0.803 (0.771–0.826)
LR+	8.79 (4.954–13.839)	3.314 (1.62–4.859)	6.731 (4.447–9.07)	6.75 (2.948–15.93)
LR−	0.148 (0.065–0.301)	0.264 (0.075–0.692)	0.144 (0.066–0.28)	0.662 (0.567–0.8)
Kappa	0.747 (0.577–0.847)	0.48 (0.177–0.641)	0.703 (0.562–0.79)	0.376 (0.213–0.495)
NMI index	0.475 (0.272–0.64)	0.223 (0.029–0.426)	0.433 (0.267–0.569)	0.125 (0.041–0.221)
Youden's J	0.768 (0.594–0.871)	0.559 (0.206–0.746)	0.745 (0.595–0.837)	0.319 (0.181–0.421)
ROC area	0.94	0.70	0.89	—

MCR, misclassification rate; OR, odds ratio; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratios for positive tests; LR−, likelihood ratios for negative tests; NMI, normalized mutual information; ROC, receiver-operating characteristic.

*Blinded for output variable in patients used as internal validation set.

Table 6. Univariate Logistic Regression Analysis Performed in the Whole Cohort of Pediatric Kidney Recipients, With Delayed Decrease of Serum Creatinine as Dependent Variable

Variable	χ^2	OR (95% CI)	EM	*P* Value
Patient continuous variables				
Patient age (y)	0.041	0.99 (0.92–1.06)	QN	.837
Height (cm)	2.72	0.98 (0.97–1.00)	QN	.112
BW (kg)	7.16	0.96 (0.94–0.99)	QN	.007
BSA (m^2)	6.7	0.23 (0.07–0.73)	QN	.013
BMI (kg/m^2)	4.52	0.93 (0.87–0.99)	QN	.044
BW1 (kg)	7.47	0.96 (0.94–0.99)	QN	.009
BWG (g/24 h)	—	0.99 —	HJ	1.000
SCr1 (mg/dL)	24.67	0.71 (0.61–0.82)	QN	<.0001
UV1 (mL/24 h)	—	0.99 —	HJ	1.000
Other continuous variables				
Donor age (y)	0.036	0.99 (0.96–1.03)	QN	.849
CIT (h)	1.76	0.93 (0.84–1.03)	QN	.183
WIT (min)	6.04	0.95 (0.91–0.99)	QN	.017
Patient categorical variables				
Sex	2.43	1.79 (0.7–4.56)	QN	.118
Pre-Tx dialysis mode	0.40	1.26 (0.59–2.7)	QN	.525
CMV	—	1.10 (0.11–10.8)	HJ	.931
EBV	—	1.10 (0.0–14.7)	HJ	.939
HCV	—	1.10 (0.11–10.88)	HJ	.931
Blood transfusions	0.32	1.10 (0.29–4.19)	HJ	.882
Diagnostic category	0.24	1.03 (0.89–1.2)	QN	.620
Other categorical variables				
Donor sex	1.98	0.57 (0.22–1.47)	QN	.159

CI, confidence interval; OR, odds ratio; EM, estimation method; QN, quasi-Newton; HJ, Hooke-Jeeves; BW, body weight on the day of kidney transplantation; BSA, body surface area; BMI, body mass index; BW1, body weight on the day 1 after kidney transplantation; BWG, body weight gain; SCr1, serum creatinine measured on the day 1 after kidney transplantation; UV1, urine volume collected in the first 24 hours after kidney transplantation; $T_{1/2(SCr)}$, the number of days needed to reach a serum creatinine level 50% below that before kidney transplantation; CIT, cold ischemia time; WIT, warm ischemia time; PD, peritoneal dialysis; HD, hemodialysis; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HCV, hepatitis C virus; DDSCr, delayed decrease of serum creatinine.

maintained a good balance between sensitivity and specificity (Table 5). In contrast, the restrictive criterion adopted in our study to define the output variable generated a lower incidence of DDSCr+ patients (27.02%) when compared to DGF+ patients (38%) as reported in a recent ANN application to cadaveric renal transplants.[21] Although a relatively low number of validation patients may affect the PPV,[24] in our internal validation set ANN overall accuracy reached 76.92% against 63% of previous ANN application to cadaveric renal transplants.[21]

The predictive performance of ANN was compared with a conventional LR model. The choice of LR was suggested for mathematical reasons: a "tanh" activation function of the logistic function is often used for the hidden units of a multilayer network,[5] as well as by previous applications of ANN technology.[8,21,24] Unlike the ANN, LR cannot be "blinded" for the output variable, requiring the whole patient cohort to generate reasonable estimates. The SCr1 was the unique variable reaching significance in a multivariate LR. Although this variable also resulted with the major sensitivity during ANN testing (Table 4), the variables that reached significance in univariate LR were mainly focused on BW-related parameters (Table 6). Conversely, the parameters selected during the test/training of ANN were more representative of the overall pattern of variables, including both donor- and patient-related parameters. Even though the LR showed an overall accuracy slightly higher than the ANN-internal validation set after blinding the network for DDSCr, it revealed a poor sensitivity and a lower Youden's J (Table 5). When the ANN model was run by entering all kidney recipients (training and validation set) and blinding the network for DDSCr, the overall accuracy was better than LR.

The ANN approach requires a new perspective for clinicians: input variables should not be viewed as potentially independent predictors but rather as part of the ANN model, reflecting the multidimensional nature of relationship between clinical parameters.[24] This study represents an effort to apply ANN technology among a cohort of pediatric kidney recipients. By adopting strict criteria for selection of input variables and definition of the output variable, our ANN model reached a more reliable predictive pattern than previous applications of ANN technology in adult kidney recipients. The availability of the source code in both Visual Basic and C++ languages may allow the development of a stand-alone ANN to validate our model in prospective studies.

REFERENCES

1. Martinek V, Lanska V, Tschernoster E, et al: The importance of early renal graft function. Nephrol Dial Transplant 8:361, 1993
2. Nicholson ML, Wheatley T, Horsburgh T, et al: The relative influence of delayed graft function and acute rejection on renal transplant survival. Transpl Int 9:415, 1996
3. Scandling J, Myers B: Pathophysiology of reduced glomerular filtration rate in delayed graft function. Curr Opin Nephrol Hypertens 6:405, 1997
4. Giral-Classe M, Hourmant M, Cantarovich D, et al: Delayed graft function of more than six days strongly decreases long-term survival of transplanted kidneys. Kidney Int 54:972, 1998
5. Bishop CM: Neural Networks for Pattern Recognition. New York: Oxford University Press; 1995
6. Cross SS, Harrison RF, Kennedy RL: Introduction to neural networks. Lancet 346:1075, 1995
7. Baxt WG: Application of artificial neural networks to clinical medicine. Lancet 346:1135, 1995
8. Furness PN, Kazi J, Levesley J, et al: A neural network approach to the diagnosis of early acute allograft rejection. Transplant Proc 31:3151, 1999
9. Shoskes DA, Ty R, Barba L, et al: Prediction of early graft function in renal transplantation using a computer neural network. Transplant Proc 30:1316, 1998
10. Van Biesen W, Vanholder R, Van Loo A, et al: Peritoneal dialysis favorably influences early graft function after renal transplantation compared to hemodialysis. Transplantation 69:508, 2000
11. Medical Policy Technology Assessment Committee review. Policy TRANS.00007H. Kidney Transplantation. Available from http://medpolicy.unicare.com. Accessed October 12, 2006

12. Mosteller RD: Simplified calculation of body surface area. N Engl J Med 317:1098, 1987

13. Hamburger R, Mattern W, Schreiber M, et al: A dialysis Modality Decision Guide based on the experience of six dialysis centers. Dial Transplant 19:66, 1990

14. Hunter A, Kennedy L, Henry J, et al: Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput Methods Programs Biomed 62:11, 2000

15. Dickson ER, Murtaugh PA, Wiesner RH, et al: Primary sclerosing cholangitis: refinement and validation of survival models. Gastroenterology 103:1893, 1992

16. Hooke R, Jeeves TA: Direct search solution of numerical and statistical problems. J Assoc Comput Mach 8:212, 1961

17. Forbes AD: Classification-algorithm evaluation: five performance measures based on confusion matrices. J Clin Monit 11:189, 1995

18. Youden WJ: Index for rating diagnostic tests. Cancer 3:32, 1950

19. Hanley JA, McNeil JB: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29, 1982

20. Ojo A, Wolfe R, Held P, et al: Delayed graft function: risk factors and implications for renal allograft survival. Transplantation 63:968, 1997

21. Brier ME, Ray PC, Klein JB: Prediction of delayed renal allograft function using an artificial neural network. Nephrol Dial Transplant 18:2655, 2003

22. Heston TF, Norman DJ, Barry JM, et al: Cardiac risk stratification in renal transplantation using a form of artificial intelligence. Am J Cardiol 79:415, 1997

23. Rodrigo E, Ruiz JC, Pinera C, et al: Creatinine reduction ratio on post-transplant day two as criterion in defining delayed graft function. Am J Transplant 4:1163, 2004

24. Banerjee R, Das A, Ghosal UC, et al: Predicting mortality in patients with cirrhosis of liver with application of neural network technology. J Gastroenterol Hepatol 18:1054, 2003