We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Forcasting the Clinical Outcome: Artificial Neural Networks or Multivariate Statistical Models?

Ahmed Akl and Mohamed A Ghoneim
Urology and Nephrology Center,
Mansoura
Egypt

1. Introduction

The field of prognostics has grown rapidly in the last decade and clinicians have been provided with numerous tools to assist with evidence-based medical decision-making. Most of these included nomograms, classification and regression tree analyses, and risk group stratification models [Grossberg JA, et al., 2006] and Artificial Neural Networks (ANN) [Djavan B, et al., 2002]. Nomograms are graphic representation of statistical model, which incorporate multiple continuous variables to predict a patient's risk of developing a specific endpoint (recurrence, survival, complications) [Kattan MW, 2005]. Each variable is assigned a scale of points according to its prognostic significance. The total score for all the variables is converted to an estimated probability of reaching the endpoint [Akl A, et al., 2008]. Statistical approaches require guesses as to how outputs functionally depend on inputs. Artificial neural networks have been used for evaluation of clinical data to provide results similar to conventional modeling methods [Freeman RV, et al., 2002]. They do not require the articulation of such a mathematical model. ANNs are complex computational systems that can provide a nonlinear approach for data analysis.

The ANNs forms a mapping from input to output nodes (simulated neurons) by extracting features from input patterns, assigning them weights, summing weights with activation functions, and propagating decisions to output nodes if activation thresholds are exceeded. Typical networks are organized into three layers of computational units (nodes) in which input/output layers are linked by hidden layers of nodes. Subject factors determine the number of input units, and the classification complexity determines the number of output units. The number of hidden units is determined by trial and error (training). Common routines start with one hidden unit and assign small arbitrary weights to all nodal connections. The network is fed sample data with known outcomes, and an error term is calculated by means of differences between known and predicted outputs. Learning consists of adjusting weights by backward pass of errors through the connections to network nodes in response to input data. Hidden units are added to achieve minimum error criteria, while constraining the number to promote generalization of input patterns and prevent over-fitting (memorization). Interconnection density determines the network's ability to correctly discriminate the outcomes. In statistical parlance, ANN models are a form of nonlinear

discriminant analysis with input units, weights, and activation functions resembling covariates, coefficients, and generalized additive models, respectively. To generate an accurate prediction model, several conditions should be fulfilled: use of a robust dataset that represents a large patient population, and the incorporation of prognostically significant variables into the model [Kattan MW, 2005]. In addition, the generated model should be validated using an independent testing group [Graefen BM, et al., 2002]. Herein, we have developed two predictive models using ANN. The first is to predict graft outcome after renal transplantation. The second is to predict patient outcome after radical cystectomy for invasive bladder carcinoma. The results of these models were compared with those of statistically-based multivariate models.

2. Artificial intelligence versus Nomogram in renal transplantation:

The importance of having a possibility to predict the graft outcome after renal transplantation does not need emphasis. This would allow the choice of the best possible kidney donor and the optimum immunosuppressive therapy for a given patient. Several methods have been used to construct a prognostic model. A multivariate analysis was used to predict the outcome of renal transplantation from deceased donor in an attempt to optimize the allocation of the recovery of organs [Poli F, et al. 2000]. In another study, multivariate analysis was used to predict creatinine levels in recipients of kidneys from living donors [Zapletal C, et al. 2004]. The probability of deceased donor-graft survival was studied using tree regression model [Goldfarb-Rumyantzev AS, et al. 2003]. ANNs were used to predict the possibility of delayed graft function after deceased donor renal transplantation [Brier ME, et al. 2003]. For the development of our ANNs, we have opted to use a feed-forward with back-propagation model since it is known for its stability and tendency not to over fit [Ripley RM, et al. 2004]. The algorithm is often described as a decision making process functioning like the human brain [Cross SS, et al. 1995]. In statistical parlance, ANN models are a form of nonlinear discriminant analysis with input units, weights, and activation functions resembling covariates, coefficients, and generalized additive models, respectively. Thus, it is not surprising that ANN applications are undermined by similar limitations and misuses afflicting conventional discriminant analysis. Schwarzer identified four frequent mistakes when applying ANNs [Schwarzer et al., 2000]:

- Overfitting models by training large, multilayer networks with small data sets.
- Neglecting traditional statistical methods due to inadequate bench marks or lack of significance testing.
- Applying naïve approaches to survival data, sometimes ignoring censorship.
- Claiming overly optimistic generalization properties.

In our study only significant univariate variables were incorporated as input units, but in multivariate cases, insignificant univariate variables sometimes become relevant confounders or effect modifiers. Since ANNs are touted as having the ability to select those items most important in performing classifications, this prior variable selection seems unnecessary [Figure 1].

Our study was designed to predict 5-year survival using cases with complete data (81 with missing data were dropped) divided into training (n=1500) and test (n=319) sets for building and validating models, respectively. We followed steps to guard against these problems. To avoid over-fitting, the ANN was restricted to one hidden layer, and the number of hidden

nodes was controlled by reasonable stopping criteria. Also, the ratio of number of observations in the training set (1500) to number of parameters in the model (361) was greater than 2, a recommended guideline [Schwarzer et al., 2000] [Figure 2, 3].
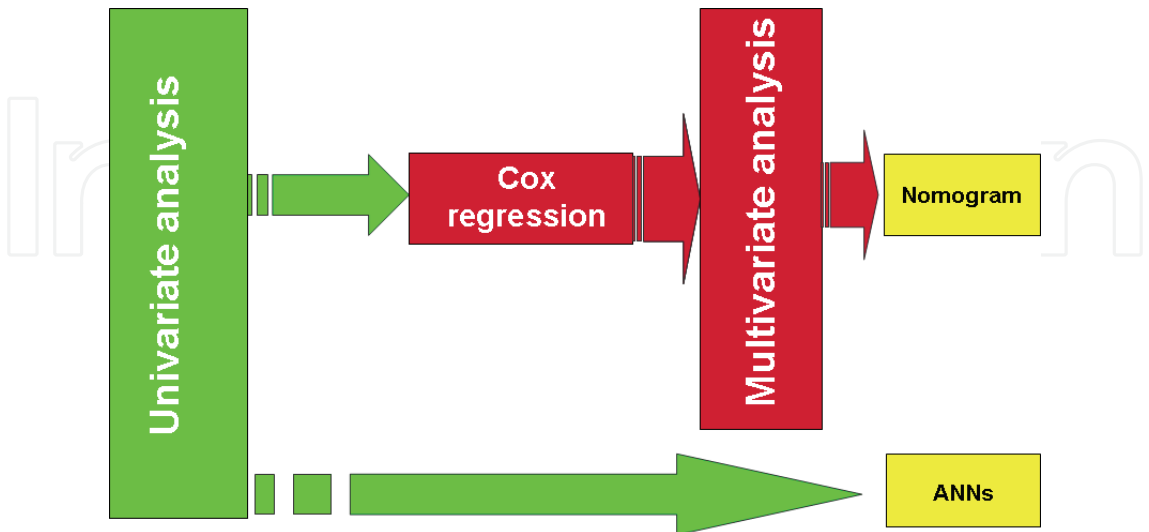
Fig. 1. Artificial neural networks model construction for predicting graft survival.
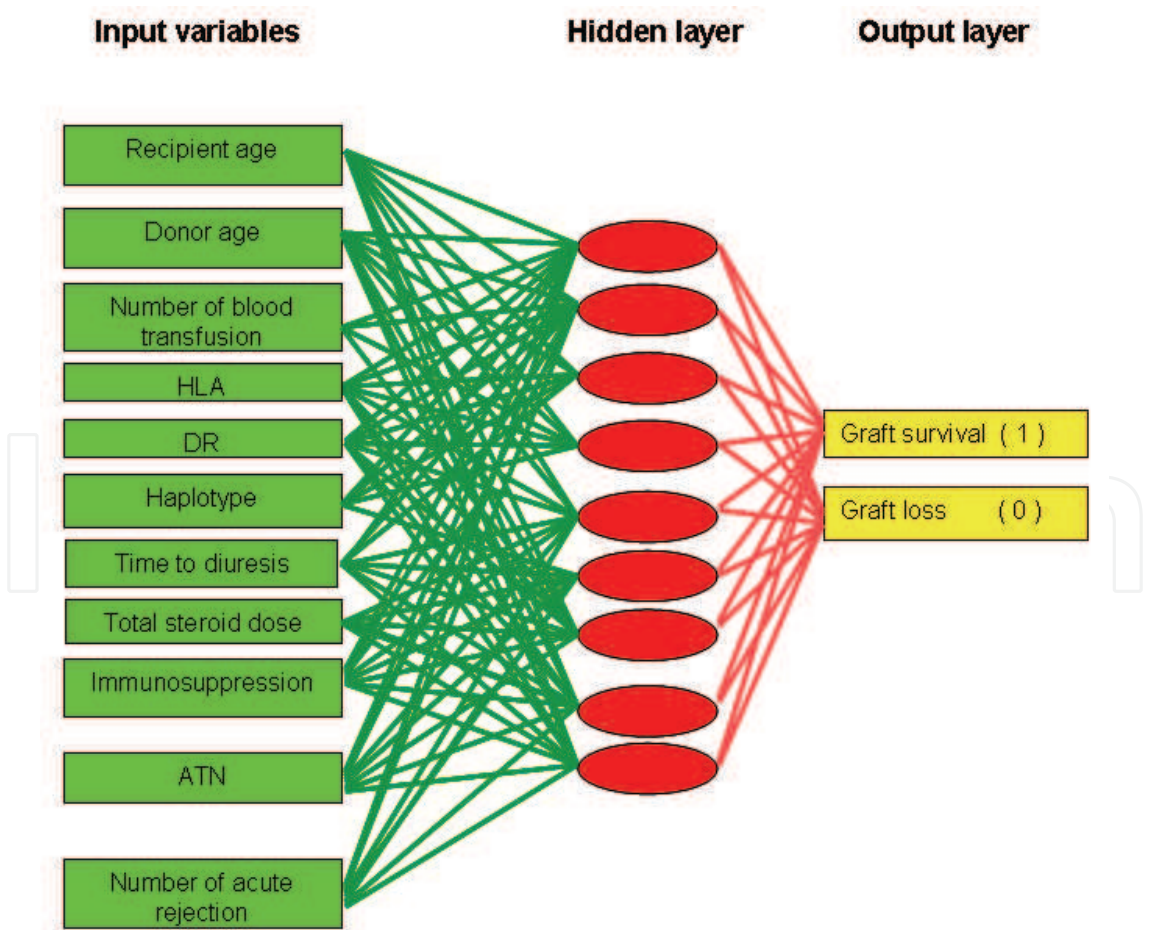
Fig. 2. Artificial neural networks architecture for predicting renal graft outcome.
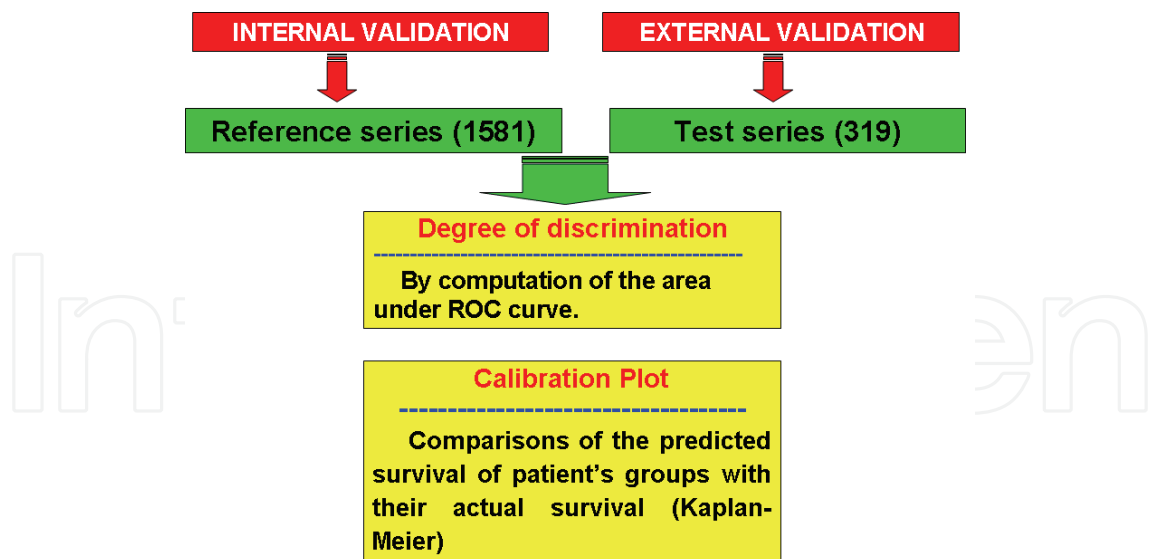
Fig. 3. Validation strategy of transplantation Ann model.

Preprocessing or normalizing data entering the ANN was done. This is an important step that usually requires a floating-point value between 0 and 1 to be assigned for each input node, with special consideration for missing values. The ANN model was tested against a nomogram built on a standard statistical approach (Cox regression) [Figure 4].
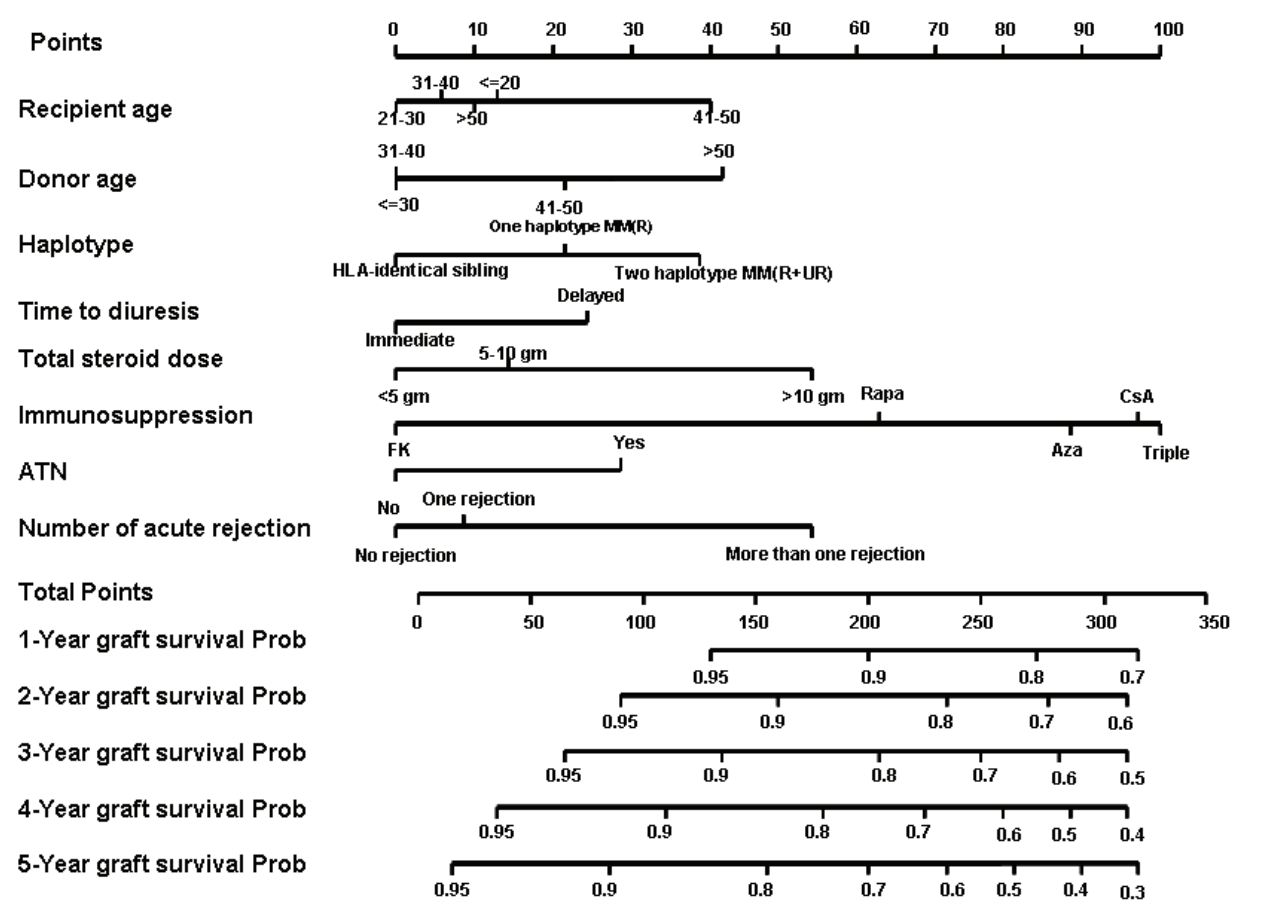

Fig. 4. Cox based nomogram for predicting renal graft outcome.

Although this nomogram was simple to interpret by clinicians, but the linear nature of Cox regression may limit its handling power of complex tasks. The nomogram was not discarded altogether but credited for easy interpretation. The imputation strategy seemed to be sufficient since calibration points between average predicted (ANN) and observed (Kaplan-Meier) 5-year survival rates for various patient groups produced virtually a 45° line. The ANNs sensitivity was 88.43 (86.4 –90.3) %, specificity was 73.26 (70 –76.3) %, and its predictions was 16% significantly more accurate than the Cox regression-based nomogram (*P*<0.001). The Cox regression-based nomogram sensitivity was 61.84% (50 –72.8) with 74.9% (69–80.2) specificity and area under ROC curve was 72% (67–77) [Table 1].

	Nomogram	ANNs
Area under ROC curve	0.72	0.88
Sensitivity at 95% specificity	0.40	0.77
Sensitivity at 90 % specificity	0.48	0.58
Specificity at 95% sensitivity	0.86	0.90
Specificity at 90% sensitivity	0.30	0.48
Hosmer and lemeshow Test X^2 (P-value)	11.6 (0.2)	14.5 (0.1)
Prediction error %	10	5

Table 1. Comparison of the predictive accuracy of renal graft survival between the Cox regression-based Nomogram and ANN.

The predicted graft survival probabilities of both models responded well to the observed graft survival. The negative predictive value for the ANNs and Cox regression-based nomogram was 82% and 86.3%, respectively. Eighty-two percent of the ANNs predicted graft survival probabilities were responding to observed graft survival, whereas only 43.5% of the Cox regression-based nomogram predicted graft survival probabilities were responding to their observed graft survival. The predictive accuracy of the ANNs prognostic model was superior to that of the nomogram in predicting 5-year graft survival. In a future project, we shall implement the ANNs model that was developed in this study in a form of software to make it available to estimate survival and prognosticate individual transplant recipients outcomes.

3. ANNs versus Nomograms and risk groups in cancer bladder:

Radical cystectomy has evolved as the primary therapeutic modality for localized or regionally advanced invasive carcinoma of the bladder. Treatment outcomes after radical cystectomy have been based on TNM staging or pathological grouping [Ghoneim M, et al. 1997]. Although this prognostic grouping yielded useful estimates of recurrence risk and survival outcomes, significant variation within each group has been observed due to the heterogeneity of the tumor biology and patient characteristics. To circumvent these limitations several mathematically based prognostic models were developed which appear to outperform clinical judgment in estimating outcome probabilities. Accurate estimates of the likelihood of success, patient outcomes and long-term morbidity are essential for patient counseling, informed decision making and to assist in clinical trial design. Clinical judgment might be biased due to subjective and objective confounders that exist at all stages of the prediction process. The

clinical TNM staging system is routinely used in decision making and patient counseling. The ease of using this tool is offset by its limited accuracy. Additional factors that can provide a predictive value, e.g. age, gender, time to surgery and other pathological features such as lymphovascular invasion, are missed. The influence of tumor stage, grade and lymph node involvement has been investigated and extensively reported [Pollack A, et al. 1995]. To improve the accuracy of pathological staging molecular markers such as P53, P21, pRb, angiogenesis and growth factor changes have been identified as having potential prognostic value [Ghoneim M, et al., 2008; Chatterjee SJ, et al., 2004]. It could be legitimately argued that the presence of a significant number of patients with bilharziasis may result in conclusions different from those of patients with cancer with typical nonbilharzial transitional cell cancer. However, on multivariate analysis the presence of bilharziasis was not an independent prognostic factor. Statistical methods based on logistic regression were traditionally used. Unfortunately complex interactions within medical data do not allow the easy use of these methods by clinicians. To overcome the shortcomings of conventional statistical methods, new models were developed to provide a means which incorporates multiple clinical, pathological properties into 1 system and achieves the best possible risk assessment. In our study the risk group stratification model identified 4 risk groups which were prospectively validated as in a previous work [Kiemeney LA, et al. 1993]. One of the potential advantages of such a system is the identification of a high risk group of patients for whom adjuvant therapy is indicated and justified [Ghoneim M, et al. 2000]. Another advantage of this model is its versatility. In the future additional prognostic variables that may influence survival significantly and independently could be included in the model once their regression estimate is determined. Construction of nomograms for survival prediction had been performed by several groups. The principle advantage of this approach is that it provides a survival probability for individual cases. Accordingly it could be maintained that they are more relevant to the patient than the group/level probabilities.

In 2006 the IBCNC published the first postoperative nomogram predicting the risk of recurrence at 5 years after radical cystectomy [Stockle M, et al. 1995]. In the same year the BCRC reported their nomogram [Bochner B, et al. 2006]. The nomogram developed by the IBCNC had a predictive index of 75% while the BCRC nomogram offered a 4% accuracy advantage. It is clear that adding more variables would improve the accuracy of any prognostic model. Shariat et al demonstrated that an abnormal profile of 2 or more abnormal markers among a panel of 5 cell cycle regulator biomarkers increased the overall predictive accuracy of the nomogram from 72.6% to 83.4% [Shariat S, et al., 2006]. Both groups demonstrated that their nomograms were more accurate and discriminating than the TNM system, solving some of the heterogeneity of outcome prediction within each stage. Nevertheless these nomograms have inherent limitations. The extent of pelvic lymph node dissection and number of lymph nodes removed were not included, although these are reported as important prognostic indices [Herr HW, et al. 2004]. In our study the operative procedure was standardized and the pathological material was revised by a single pathologist. The positive predictive index was 82.9%, which is superior to that of previously mentioned studies. Artificial neural networks are complex computational systems that can perform a large number of mental tasks. They have many advantages compared to models based on logistic regression in that they do not require statistical training, they can deal with complex nonlinear relationships and detect possible interactions among predictor variables [Tu.JV, 1996]. In a recent investigation Bassi et al compared the prognostic accuracy of ANNs and logistic regression analysis for patients undergoing radical cystectomy [Bassi P,

et al. 2009]. They concluded that the ANNs accurately predicted survival and had a prognostic performance comparable to that of the logistic regression. In our study the 3 models of risk group stratification, nomogram and ANNs were compared to assess accuracy and performance characteristics.

4. Study layout

Significant variables by multivariate analysis were used for building the risk group stratification model as well as for construction of the nomogram. On the other hand all the studied factors were entered directly into the ANNs [Figure 5, 6].
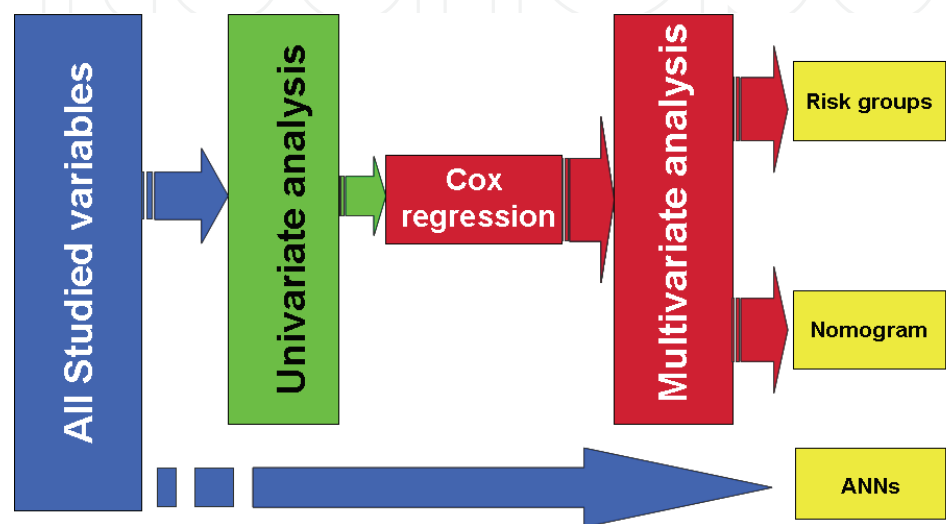
Fig. 5. Artificial neural networks model construction for predicting patient survival after radical cystectomy.
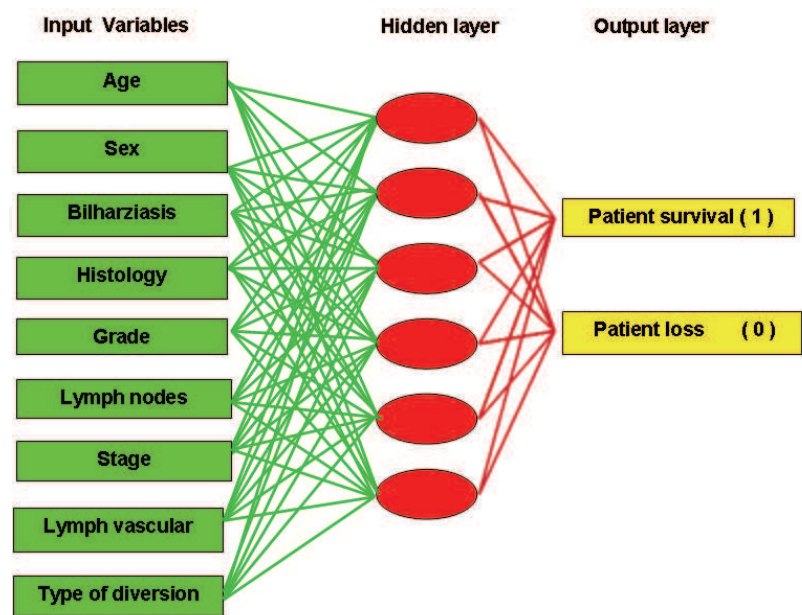

Fig. 6. Artificial neural networks architecture for predicting patient survival after radical cystectomy.

5. Risk group stratification

5.1 Model: construction and validation

The regression coefficient (B) of the significant factors on multivariate analysis was used to construct the model for risk group stratification. The algebraic sum of the regression coefficients for each case represented a proportional hazard score (Y). The possible range of Y values of patients in the reference series was divided equally into 4 groups, whereby those with the lowest Y score represented the lowest risk category and those with the highest Y score the highest risk category. Survival curves were then constructed for each of the 4 groups using the Kaplan-Meier method. Patients in the test series were classified into 4 groups according to the previously mentioned proportional hazard score (Y). For external validation survival curves were constructed for each risk category of the test series and compared with the corresponding curves of the reference series. Statistical differences were determined using the log rank test with p <0.05 considered significant. In addition, the degree of discrimination was determined by computing the ROC curve [Figure 7].

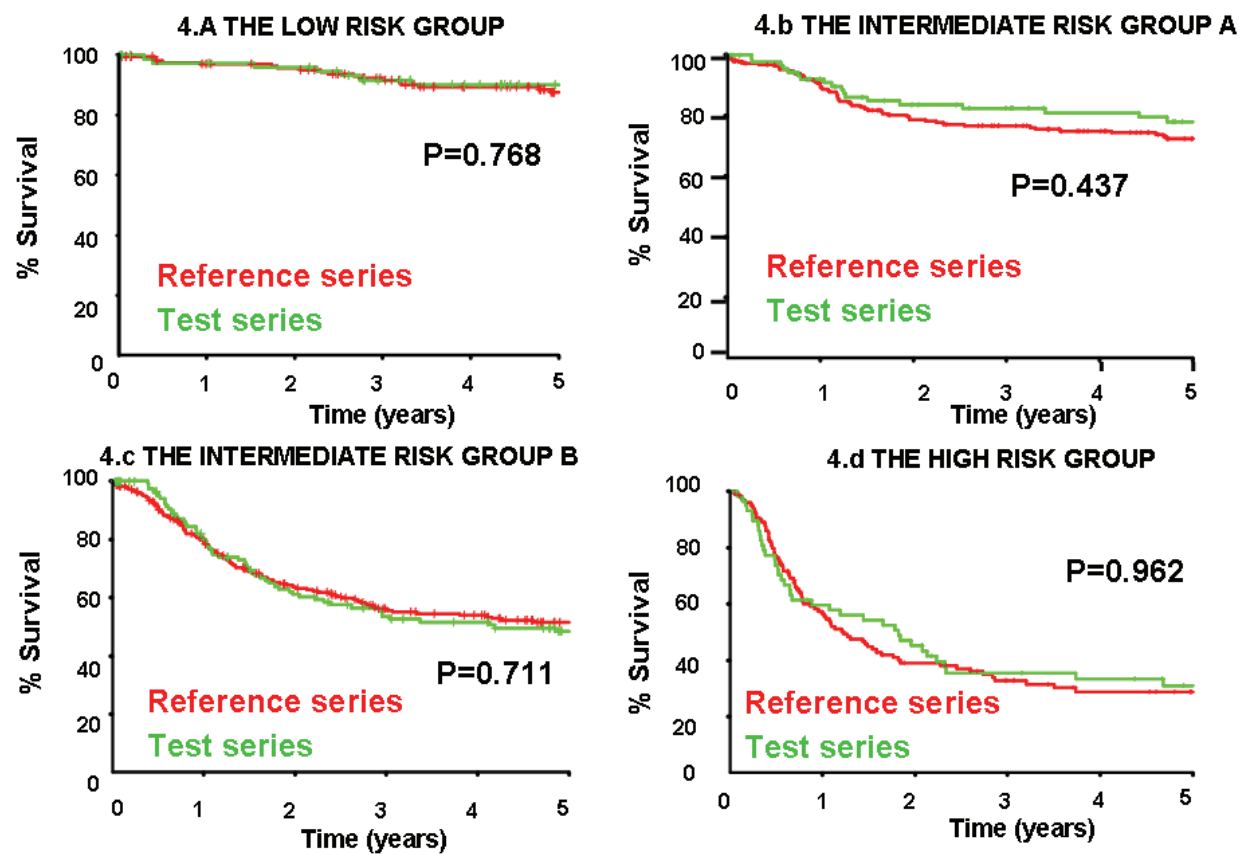Fig. 7. External validation of risk grouping for predicting patient survival after radical cystectomy.

5.2 Nomogram: construction and validation

For construction of the nomogram the R software package version 2.7.0 was used (R foundation for statistical computing, Bell Laboratories, Lucent technologies, with Frank Harrel's Design and Hmisc libraries included). Each variable was assigned a scale of points according to prognostic significance which ranged from 0 to 100. The point values

determined for each individual case were added to give a total sum. The total sum thus calculated was correlated to the 5-year patient survival probability of the same case. Validation of the nomogram involved internal as well as external validation. Internal validation was performed by discrimination as well as calibration. The degree of discrimination was quantified by computation of the area under the ROC curve. The calibration was assessed by grouping patients with respect to nomogram predicted survival probabilities. The group means were then compared with the observed Kaplan-Meier estimates. For the discrimination and calibration steps a total of 200 bootstrap re-sampling was used to obtain less biased estimates. The test series was used for external validation. The degree of discrimination was again achieved by determination of the area under the ROC curve [Figure 8].
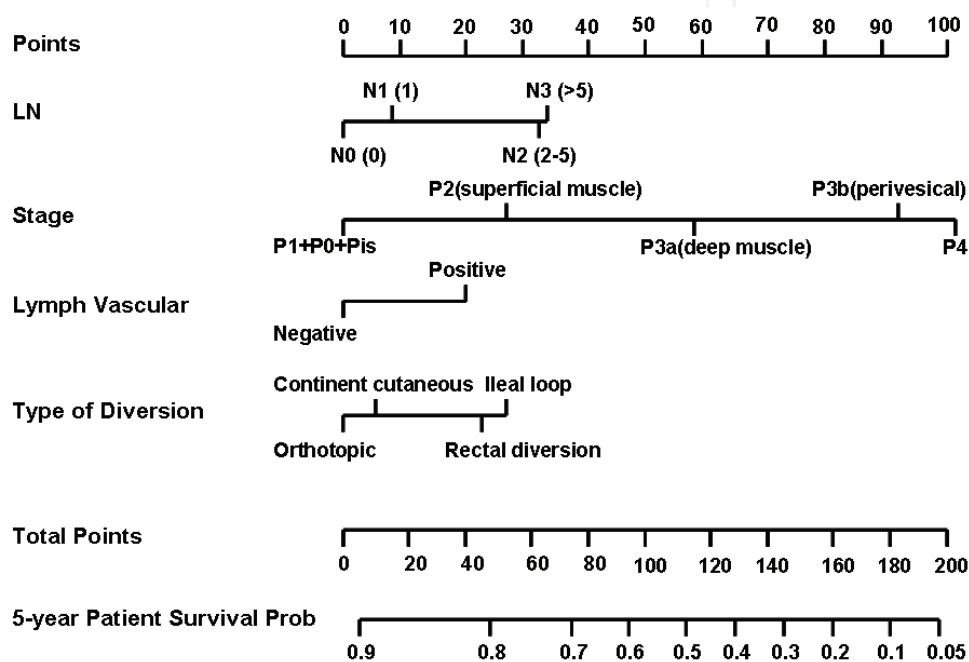

Fig. 8. Cox based nomogram for predicting patient survival after radical cystectomy.

5.3 ANNs: Construction and validation

All studied variables were used as input variables for ANN construction. The model used is commercially available software (SPSS version 16). The basis of this network is multilayer perceptrons that have feed-forward back-propagation topology. The network consists of 3 layers.

1. The input layer is of 31 neurons into which variables were entered following their normalization through a process process of standard rescaling (subtract the mean and divide by the standard deviation).
2. A hidden layer is composed of 6 neurons in which computation and differential weighing of the different variables are performed.
3. The output layer is 2 neurons into which the end results of therapy are entered (patient survival or loss).

Training was performed by a conjugate gradient descent optimization algorithm. The target was to achieve a sum of squared error of 0.01. The model starts with 1 neuron in the hidden layer, and when that neuron fails to improve the square error at the end of the training cycle

another neuron is automatically added to the hidden layer to improve the model performance. This process is terminated once a sum of square error of 0.01 has been achieved or 1,000 training cycles are completed. Six neurons in the hidden layer and 200 epochs (training cycles) were necessary to achieve this goal. After training the system was tested by internal and external validation using the same methodology as that used for the nomogram.

6. Comparison of risk group

Stratification, nomogram and ANNs

The ability to discriminate between patient survival or loss using the 3 models was compared by estimation of the area under the ROC curves (MedCalc® software version 9.3.7.0) Using these curves the sensitivity and specificity of the 3 models were also obtained at certain cutoff points. Comparisons between the performance and accuracy of the three models are provided in Figure 8 and Table 4. The ANNs outperformed the risk group stratification and nomogram models in predicting 5-year patient survival probabilities. ANNs were more sensitive and specific with a larger area under ROC curve than the other two models. The area under ROC curve for the ANNs was 0.86 while it was 0.72 for the risk groups and 0.74 for nomogram. Further more the ANNs demonstrated a superior positive as well as negative predictive values [Table 2].

	Risk groups (A-D)	Nomogram	ANNs
Area under ROC curve (95% CI)*	0.72 (0.67-0.77)	0.74 (0.69-0.78)	0.86 (0.82-0.89)
Sensitivity (95% CI)*	61.06 (54.4 - 67.5)	60.44 (53.7 - 66.9)	78.7 (72.7-83.9)
Specificity(95% CI)*	75.38 (67.1 - 82.5)	78.46 (70.4 - 85.2)	81.25 (73.4-87.6)
Positive LR#	2.48	2.81	4.20
Negative LR#	0.52	0.50	0.26
Positive predictive value	81.2	82.9	87.9
Negative predictive value	52.7	53.4	68.9
Hosmer and lemeshow Test X2 (P-value)	0.16 (0.921)	10.7 (0.219)	13 (0.112)

Table 2. Comparison of the predictive accuracy of patients survival following radical cystectomy between Risk groups, nomogram and ANNs.

Predictive models for invasive bladder carcinoma can accurately stratify patients according to risk of recurrence, progression and treatment tolerability. This enables the clinicians to make the best decision in patient counseling, selecting the optimal adjuvant therapy and follow up schedule. Among the available models our comparative study provides evidence that ANNs have several advantages compared to risk groups and nomogram models, the most important being the higher predictive accuracy, better performance and clinical applicability. However, risk groups and nomograms rely on methodologically sound and valid alternatives that cannot be ignored. Despite their advantages predictive tools cannot replace clinical judgment. Their input has to be weighed against several other considerations

such as comorbidity, cost, and social and psychic factors. Finally improvements in the current models can be attained by incorporating modern imaging tools and novel biomarkers. Moreover validation in large patient cohorts and prospective data acquisition need no emphasis.

7. Conclusions

The Artificial neural networks outperformed the risk groups and nomogram models in predicting the 5-year patient survival. Furthermore, ANN have a more satisfactory prognostic performance in terms of specificity and sensitivity compared to the other models.

8. References

Akl A, Mostafa A, Ghoneim MA. (2008). Nomogram that predicts graft survival probability following living-donor kidney transplant. *Exp Clin Transplant*; 6: 30.

Bassi P, Sacco E, DeMarco V, Aragona M, Volpe A. (2007). Prognostic accuracy of an artificial neural network in patients undergoing radical cystectomy for bladder cancer: a comparison with logistic regression analysis. *BJU*; 99 : 1007 – 12.

Bochner B, Kattan M, Vora K. (2006). Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. *J Clin Oncol*; 24 (24): 3967 – 72.

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

Chatterjee SJ, Datar R, Youssefzadeh D, George B, Goebell PJ. (2004). Combined effects of p53, p21, and pRb expression in the progression of bladder transitional cell carcinoma. *J Clin Oncol*; 22 (6) : 1007 – 13.

Cross SS, Harrison RF, Kennedy RL. (1995). Introduction to neural networks. *Lancet*; 346: 1075.

Djavan B, Remzi M, Zlotta A, et al. (2 002).Novel artificial neural network for early detection of prostate cancer. *J Clin Oncol*; 20: 921.

Freeman RV, Eagle KA, Bates ER, et al. (2002). Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty. *Am Heart J*; 140: 511.

Ghoneim M, Abdel-Latif M, El-Mekresh M, Abol-Enein H, Mosbah A. (2008). Radical cystectomy for carcinoma of the bladder: 2720 consecutive cases, 5 years later. *J Urol*; 180 : 121 – 7.

Ghoneim M, El-Mekresh M, Mokhtar A, Gomha M, El-Baz M, El-Attar I.(2000). A predictive model of survival after radical cystectomy for carcinoma of the bladder. *BJU*; 85 : 811 – 6.

Ghoneim M, El-Mekresh M, El-Baz M, El-Attar I, Ashamallah A. (1997). Radical cystectomy for carcinoma of the bladder: critical evaluation of the results in 1026 cases. *J Urol*; 158 : 393 – 9.

Goldfarb-Rumyantzev AS, Scandling JD, Pappas L, et al. (2003). Prediction of 3-year cadaveric graft survival based on pre-transplant variables in a large national dataset. *Clin Transplant*;17: 485.

Graefen BM, Karakiewicz PI, Cagiannos I, et al. (2002). Validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer. *J Clin Oncol*; 20: 951.

Grossberg JA, Reinert SE, Monaco AP, et al. (2006). Utility of a mathematical nomogram to predict delayed graft function: A single-center experience. *Transplantation*; 81: 155.

Herr HW, Faulkner JR, Grossman HB, Natale RB, deVere White R, (2004). Surgical factors influence bladder cancer outcomes. A cooperative group report. *J Clin Oncol*; 22 : 2781 – 9.

Kattan MW.(2005). When and how to use informatics tools in caring for urologic patients. *Nat Clin Pract Urol*; 2: 183.

Kiemeney LA, Witjes JA, Verbeek AL, Heijbroek RP, Debruyne FM.(1993). The clinical epidemiology of superficial bladder cancer. Dutch South-East cooperative urological group. *B J Cancer* ; 67 (4): 806 – 12.

Poli F, Scalamogna M, Cardillo M, et al. (2000). An algorithm for cadaver kidney allocation based on a multivariate analysis of factors impacting on cadaver kidney graft survival and function. *Transpl Int*; 13(suppl 1): S259.

Pollack A, Zagars G, Cole C, Dinney C, Swanson D, (1995). The relationship of local control to distant metastases in muscle invasive bladder cancer. J Urol; 154 : 2059 – 64.

Ripley RM, Harris AL, Tarassenko L. (2004). Non-linear survival analysis using neural networks. *Stat Med*; 23: 825.

Schwarzer G, Vach W, Schumacher M. (2000). On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med*; 19: 541.

Shariat S, Karakiewicz P, Palapattu GS, Amiel GE, Lotan Y, et al. (2006). Nomograms provide improved accuracy for predicting survival after radical cystectomy. *Clin Cancer Res*; 12 : 6663 – 76.

Stockle M, Meyenburg W, Wallek S, Voges GE, Rossmann M, et al. (1995). Gertenbach U, Thurof JW, Huber C, Hohenfellner R. Adjuvant poly-chemotherapy of non-organ-confined bladder cancer after radical cystectomy. *J Urol*; 153 : 47-52.

Tu.JV . (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* ; 49: 1225-31.

Zapletal C, Lorenz MW, Woeste G, et al. (2004). Predicting creatinine clearance by a simple formula following live- donor kidney transplantation. *Transpl Int*; 17: 490.
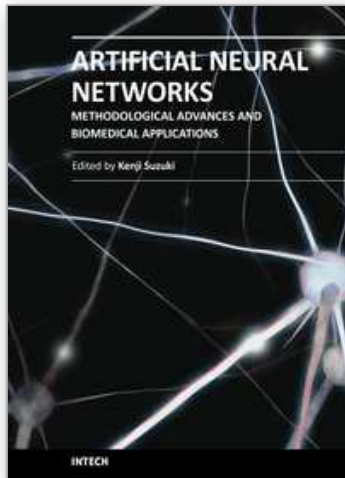
Artificial Neural Networks - Methodological Advances and Biomedical Applications

Edited by Prof. Kenji Suzuki

Artificial neural networks may probably be the single most successful technology in the last two decades which has been widely used in a large variety of applications in various areas. The purpose of this book is to provide recent advances of artificial neural networks in biomedical applications. The book begins with fundamentals of artificial neural networks, which cover an introduction, design, and optimization. Advanced architectures for biomedical applications, which offer improved performance and desirable properties, follow. Parts continue with biological applications such as gene, plant biology, and stem cell, medical applications such as skin diseases, sclerosis, anesthesia, and physiotherapy, and clinical and other applications such as clinical outcome, telecare, and pre-med student failure prediction. Thus, this book will be a fundamental source of recent advances and applications of artificial neural networks in biomedical areas. The target audience includes professors and students in engineering and medical schools, researchers and engineers in biomedical industries, medical doctors, and healthcare professionals.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ahmed Akl and Mohamed A Ghoneim (2011). Forcasting the Clinical Outcome: Artificial Neural Networks or Multivariate Statistical Models?, Artificial Neural Networks - Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978-953-307-243-2, InTech, Available from: http://www.intechopen.com/books/artificial-neural-networks-methodological-advances-and-biomedical-applications/forcasting-the-clinical-outcome-artificial-neural-networks-or-multivariate-statistical-models-

INTECH

open science | open minds