ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

Artificial neural networks and prognosis in medicine. Survival analysis in breast cancer patients

Leonardo Franco[1,2], Jose M. Jerez[2], Emilio Alba[3]

[1]Department of Experimental Psychology
University of Oxford
South Parks Road, OX1 3UD
Oxford, United Kingdom

[2]Departamento de Lenguajes y Ciencias de la Computación
Universidad de Málaga
Campus Universitario de Teatinos, s/n
29071 Málaga, Spain

[3]Servicio de Oncología
Hospital Clínico Universitario Virgen de la Victoria
Campus Universitario de Teatinos, s/n
29071 Málaga, Spain

Abstract. In this paper we first give an introduction to the problem of prognosis in medicine. The importance of prognosis is highlighted and a brief summary of some successful applications of neural networks is included, together with an analysis of their advantages over the standard statistical tools. In the second part, we compare the performances of Cox proportional hazard model and an approach based on artificial neural networks constructed for the prognosis of outcome in patients with primary breast cancer. The data was collected from 32 hospitals in Spain, via the Spanish group of research in breast cancer within the framework of the "El Alamo" project. The population was divided into training and test sets, and the predictive accuracy of the prognosis models (Cox and neural networks) was compared by determining sensitivities, specificities and the area under receiver operating characteristic curves (area ROC). The results show that neural network predictions are much more accurate, in particular in the early months after surgical intervention.

1 Introduction to Prognosis in Medicine and Neural Networks

Prognosis can be defined as an estimation of the probable course of a disease in a particular patient. It has been always considered an important part of the medicine process but in general it has attracted less attention than diagnosis. Prognosis plays a key role in medical decision making, personal decision making, health policy and medical research [16] .

The choice between different treatments is made by comparing the prognoses associated with the respective treatments and while in the past doctors often took decision on behalf of the patients, nowadays patients are increasingly

involved and base their decision also on their personal values. From a health policy perspective prognosis is quite an useful tool as it provides a comparison for the different treatments and factors influencing it, and thus is a key element in the task towards optimization of resources. Prognosis in medicine remains an essential element in modern medical practice as it meets patients needs for information about their future and provide a rational for taking medical decisions. Standard statistical methods have been partially successful at estimating individual prognosis and in recent years the introduction and use of artificial neural networks (NNs) have brought a new perspective in the field [16, 4]

1.1 Standard statistical methods for survival analysis

In prognostic studies in general, the primary outcome of interest is time to an event such as relapse, progression, death, etc., and this analysis is termed survival analysis. A survival time distribution can be characterized by its survival function $S(t)$ defined as the probability of survival at least until time t. An alternate but related characterization is through the hazard function $\lambda(t)$ which is the instantaneous rate of failure at time t. The survival function $S(t)$ and the hazard function $\lambda(t)$ are totally related and thus the specification of anyone of them is enough to determine completely the other.

The Cox regression model, the standard statistical tool in survival analysis, is a semiparametric regression model. It specifies a model for the hazard function having the following form:

$$\lambda(t) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_1 x_1 + \ldots + \beta_n x_n)$$

The function $\lambda_0(t)$ is called the baseline hazard function and includes all the time dependency in the model. The x_i are the explanatory variables or **c**ovariates, like tumor size, age of the patient, etc. and β_i are the coefficients for each covariate. It has the advantage of permitting a simple interpretation of the β coefficients: a unit increase in a covariate x_i corresponds to multiplication of the baseline function by the factor $\exp(\beta_i)$ (if all other covariates are held fixed). The disadvantage is that, as said before, all the time dependency of the model is included in the baseline function and thus this implies that the effect of the covariates is constant in time. Whether this is true from a biological perspective is an unknown fact in most cases and for the application and validity of the Cox model an assessment of the proportional hazard assumption should be made.

One particular feature of the data used for prognosis analysis is the existence of censored data, cases in which the state of the patient after certain point in time (before the event of interest has occurred) is unknown. Unfortunately, excluding these cases significantly biases the results and thus can not be excluded. The Cox proportional hazard analysis [8] is an accepted solution to the problem of analysing censored data and has become the standard survival analysis statistical technique. However, the Cox model is mainly used to study the importance of covariates for survival but it is seldom used to estimate survival times [26, 31].

In order to make predictions of survival for new cases the value of the β coefficients have to be estimated but also the baseline hazard function has to be

computed. The estimation of the hazard baseline function is a non-trivial task and a wrong baseline choice can dramatically affect the results [26].

1.2 Neural Networks in survival analysis

Two aspects make the neural networks more powerful than Cox regression analysis. One is that the effect of the covariates can be made time dependent by the introduction of the time as input to the network (see section 2.2.1) and second that the effect of the covariates might be non-linear and thus these effects can be captured by the neural network. The main disadvantage of the NNs models is that they allow little insight on which variables are most influential in the model and also about the precise relationship between outcome and explanatory variables.

Neural networks are quite flexible in the sense that they do not need a baseline hazard function. Different approaches can be taken to predict prognosis using NN, as they can be used to model the hazard function, and also the probability of survival at fixed or different times. NNs application require the selection of an architecture and also of a training algorithm. In general, most of the architectures used contains a single hidden layer with a number of units between 3 and 50, according to the complexity of the problem and the chosen architecture is normally selected by trial-and-error method. Regarding training, Backpropagation is the standard training algorithm used but in general needs to be combined with some method to prevent overfitting, as weight-decay or early stopping [21].

The application of NNs to survival prediction has been in general quite successful, normally outperforming the Cox model and thus there have been many studies reporting their use in medicine prognosis. The previous fact, indicates the existence of non-trivial non-linear relationships between the covariates that can not be easily captured by parametric models [26, 6, 31].

Several neural network approaches have been proposed to model survival data. In some cases (see for example [12, 28, 22, 5, 25]) the prognostic covariates have been used as inputs to the neural system while the time to relapse is the output of the neural network. These previous approaches implement a separation between the dependence on time and on the patient data resulting in non-linear proportional hazards models. A more efficient representation of time is to include it as a covariate, and in this case the output of the system becomes an indicator of relapse or not at a given time. This kind of approach (also refereed as time-coded models) has been implemented by several authors [10, 9, 24, 27, 1, 3] and can be interpreted as the discrete time implementation of the proportional hazards model [26]. Moreover, this kind of neural network model for survival prediction has proved to be very stable in monthly studies over follow-up periods of several years [2]. Time-coded models generate a prognostic index that can be interpreted as conditional probabilities or cumulative probabilities depending on the preprocessing performed on the input data [11].

2 Case Study: Survival analysis applied to breast cancer patients

Decisions as how to treat **breast cancer** patients after surgery have been contingent on the accuracy of estimating the behaviour and outcome of the disease. Histological, biochemical and clinical information have demonstrated to have prognostic utility, however predicting the disease outcome for an individual patient remains a challenging task.

Using a large database containing data from 3811 patients cases obtained from 32 different hospitals across Spain (via GEICAM, the Spanish Group in Breast Cancer Research), we implemented neural networks architectures to predict the probability of breast cancer relapse and we compared the performance of the neural networks to the obtained using the standard Cox survival model. We have also carried a detailed analysis of the importance of including (or not) different clinical markers as inputs in the neural architectures. The analysis was carried both for the whole dataset but also limited periods of times were considered in order to assess the time dependency of the prognostic covariates and as a way to improve the accuracy of the system by selecting on each interval the most appropriate variables. It is important to outline that real data sets with significant large numbers of cases have not been extensively analyzed using neural networks as most of the works reported in the literature use small data sets [11].

2.1 The data set

Data were collected from the "El Alamo" Project, the largest database on breast cancer in Spain. The dataset analyzed in this study includes demographics, therapeutic and recurrence-survival information from 3811 women patients with operable invasive breast cancer diagnosed in 32 different hospitals belonging to the Spanish Breast Cancer Research Group (GEICAM) between the years 1990 and 1993. All the patients were characterized for a set of clinical and pathological variables specified in Table 1. The analysis was restricted to patients with follow-up time of at least one month, and thirty-four percent of patients were relapsed in the period of study. The median follow-up (i.e. the time elapsed from the date of surgery to the last updating of the patient record) was 76 months (1 - 128 months).

2.2 Prognostic models

2.2.1 Artificial neural networks

Feed-forward neural networks can be seen as analogous to regression models, in which covariates are called inputs, coefficients are called weights, and the outcome variable is called output. In this work, a three-layer neural network model (an input layer, with each input node corresponding to a prognostic factor plus one node for the coded time; a hidden layer; and an output layer) was constructed with an ad-hoc software developed in C++ and R languages

Prognostic variables (mnemonic)	Range	Mean	SD	Median
Age, years (A)	$25-90$	56.21	12.40	57
Tumour size (T)	$0.2-13$	2.87	1.64	2.5
No. Axillary lymph nodes (N)	$0-35$	2.48	4.23	1
Histological grade (G)	$1,2,3$	2.04	0.65	NA
Histological type (H)	$1,2,3$	1.19	0.53	NA
Hormonal receptor status (ER)	$1,2$	1.72	0.44	NA
Menopausal state (M)	$1,2$	1.67	0.47	NA
Type of treatment (Tr)	$0,1,2,3,7,$ $8,9,10,11$	5.37	3.36	NA

Table 1: Summary of patient data: range, mean, SD and median

and backpropagation was the training algorithm When the desired output takes only two values, as in the case considered in this work (relapse or not relapse) the cross-entropy error function has demonstrated to have a better performance [15]. Transfer functions for all neurons in the network were sigmoidal and overfitting problems were avoided using a regularization technique known as weight decay [15]. Weight decay is used to prevent the synaptic weights from excessive growing as this has been demonstrated to improve the generalization ability [Bartlett, 1997]. The number of neurons in the hidden layer was determined using a constructive process, in which we consider different architectures with a number of neurons ranging from 5 to 30 neurons in the hidden layer. We did not consider larger networks as this led to no further improvement in network performance. Sixteen different combinations of covariates were analyzed in order to identify the best set of prognostic factors in terms of the accuracy in the prediction using a validation set. The variables age (A), tumor size (T), number of axillary lymph nodes (N) and grade of tumor (G), considered as very significant prognostic factors in clinical standard practice [13, 23, 7] were incorporated to every dataset; whereas menopausal status (M), histological type (H), estrogen receptors (ER) and type of treatment (Tr) were included (or not) in the sixteen combinations considered. The input data to be fed into the neural architecture was pre-processed by Gaussian normalization that makes the data to be normally distributed around 0.5 with standard deviation equals to 1. Data containing missing values were not considered and outliers with covariate values 3 standard deviations larger than the mean were also eliminated from the dataset. For many of the intervals considered and for the general case, we used the following 5 covariates: A, T, N, G and Tr (see Table 3).

The neural approach adopted in this work lies within those known as time-coded models, in which the time of follow-up is included as an additional covariate. The input vectors are replicated from the first time interval until the interval previous to the maximum follow-up, setting the survival status to 0 and the time of follow-up to the mean value of the corresponding interval (5 months, 15 months, etc). Besides, for a patient who has died, data vectors are included with survival status 1 for all time intervals after the occurrence of the event. The selective replication of cases for all the patients depending on the censoring sta-

tus at maximum follow-up makes the output of the network to represent directly the cumulative relapse probability for a given patient, and it has also the advantage that a single neural network can be used to obtain predictions for every time interval. In Table 2 examples illustrating the way in which the data from censored and non-censored patients was fed into the neural network are shown. The patient case shown in the first row of Table 2 is an example of a censored patient with a last follow-up at 53 months. This kind of right-censored data was replicated in all the intervals preceding the last follow-up setting a negative value for the outcome, and for the intervals occurring after the time of the last follow-up the data was no longer considered (as in Kaplan-Meier analysis). The example in the last row of Table 2 corresponds to a patient who has relapsed at 27 months and for which the data was replicated setting a negative (using a numerical value of 0) outcome for the preceding intervals to the time of relapse while the outcome was set to 1 (positive outcome) for the later intervals (this was done to make the output of the network to determine the cumulative probability of relapse). The time used as input for each of the intervals considered was the mean value of the interval (5, 15, 25, 35, 45, 55 and 93) except when the precise value of the time of relapse or censorship was known.

	$0-10$	$11-20$	$21-30$	$31-40$	$41-50$	$51-60$	61−MFT
Output	0	0	0	0	0	0	–
Input (months)	5	15	25	35	45	53	–
Output	0	0	1	1	1	1	1
Input (months)	5	15	27	35	45	53	93

Table 2: Two examples showing how the data from the patients was fed into the neural network in the different time intervals considered. The patient in the second row is an example of a censored patient with a last follow-up at 53 months. The example in the last row corresponds to a patient who has relapsed at 27 months and for which the data was entered with negative outcome for the preceding intervals to the relapse while was considered with positive outcome for the later intervals

2.2.2 Cox regression analysis

Cox regression analysis, a standard statistical tool in survival analysis, was used as a comparison to the neural network approach performance. The relationships between different prognostic factors and patient survival, as well as the calculation of the prediction of the patient outcome, were assessed using Cox proportional hazards regression [8] using the COXPH and PREDICT procedures in R [14, 30]. Variable selection was done using backward and forward stepwise selection processes (the significance level of entry and permanence of a given variable in the model was $p < 0.05$). Tied event times were handled by the Breslow method and estimation of the survivors functions at event times were performed using the BASEHAZ statement in R. The assumption of hazard proportionality for the model was tested using the ZPH procedure in R, which performs a test for a non-slope in a generalized linear regression of the scaled

Schoenfeld residuals on functions of time [14]. A p value < 0.05 in the test for zero-slope, or a non-constant value for the parameters b over time when the scaled Schoenfeld residuals are plotted, indicated a violation of the proportional hazard assumptions.

2.2.3 Validation of the prognostic models

In order to estimate the classification accuracy for both neural network and Cox regression models, a standard technique of stratified nine-fold cross-validation was used [29]. Firstly, data was divided into 10 subsets of approximately equal size, and one of them was reserved to test the prediction accuracy for every prognostic model. Next, each of the left 9 random subsets of the data served as a validation set (to select an appropriate neural architecture) for the prognostic model estimated. Then, the prediction accuracy for the models was tested over the test data subset. Finally, the overall prediction accuracy for the model is then assessed as an average over 9 experiments. For both models (Cox and neural networks), a survival curve for each patient in the test set was generated, obtaining the cumulative probability of survival for every time interval. The predictive accuracy of the different models was computed using the area under the ROC curve for censored data [19, 20]. The ROC area (AUC) is an expression of the probability that a randomly drawn individual from the positive reference sample has a greater test value than a randomly drawn individual from the negative reference sample [17, 28]. Areas under the ROC curves (AUCs) of different models were compared by the Hanley-McNeil procedure [18, 17].

We computed the prediction accuracy of both neural network and Cox regression models in two different ways. First, all the available data was analyzed together, and second, different ANN models were trained to predict the probability of relapse in different time intervals, so different neural architectures can be used for each time interval considered. By using the cross-validation architecture selection procedure (see section methods) different neural network architectures with a single hidden layer with a number of neurons between 5 and 30 were chosen. The networks were trained by backpropagation with learning rate $\varepsilon = 0.05$, combined with a weight decay procedure with parameter λ chosen within the range $[1.0e-5, 7.5e-1]$ as the value for which the validation error was the lowest of all the tested architectures.

2.3 Simulation results

When using the artificial neural network model, the best performance was obtained with five input variables that were age, tumor size, number of affected axillary lymph nodes, grade of tumor and type of treatment, when all data were considered together independently of the time interval. The area under the ROC curve (AUC) was 0.8497 (SE = 0.015, 95% CI 0.82-0.87) and a graph of the curve is plotted in Figure 1.

When Cox regression model was applied the variables that were statistically significant applying both forward and backward stepwise selection procedures

were age (OR = 0.988, 95% CI 0.98-0.99, $p = 0.001$), tumour size (OR =1.135, 95% CI 1.09-1.18, $p < 0.001$), axillary lymph nodes (OR = 1.082, 95% CI 1.07-1.10, $p < 0.001$), grade (OR = 1.399, 95% CI 1.24-1.58, $p < 0.001$), histological type (OR = 0.794, 95% CI 0.68-0.93, $p = 0.004$), type of treatment (OR = 0.957, 95% CI 0.93-0.98, $p = 0.002$). Hormonal receptors and menopausal status (both with $p > 0.05$) were excluded from the full model and no interactions were found between any of the variables. The area of the ROC curve for the Cox regression model with the same generalization dataset used for the neural network was 0.7669 (SE = 0.0178, 95% CI 0.74-0.79). The difference between the two ROC was highly statistically significant ($z - score = 3.5276, p < 0.001$).

As mentioned above, we also constructed specialized neural architectures to compute predictions of survival for 7 different time intervals using neural architectures with a single hidden layer containing between 5 to 30 neurons in all cases. The results are shown in table 3, where the intervals considered are indicated in the first column and the results for the Cox and neural networks models are shown in columns 2 and 3 respectively (mean value $\pm$ standard error). For the case of the neural network the different prognostic factors used in each interval are shown in column 3 and in the last column of table 3 the values of the $z - scores$ (and related one tail $p - values$) of the difference between the predictions of neural network (ANN) and Cox regression models are included.

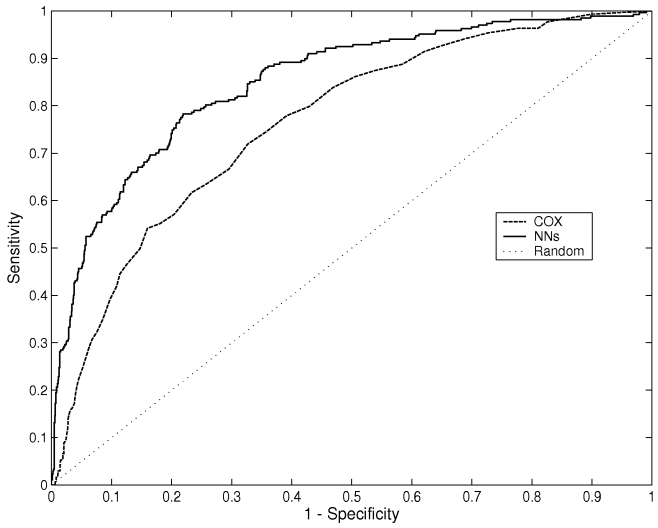

Figure 1: ROC curves obtained by using both Cox and neural networks models for the prediction of breast cancer relapse. The areas under the ROC curve (AUC) were 0.8497 and 0.7669 for the NNs and Cox model respectively, and the difference was statistically significant at $p < 0.001$.

From these results, we can argue that using a single neural network constructed to predict the prognosis for all time intervals together (from one month

Survival period (months)	Cox $AUC \pm SE$	ANN $AUC \pm SE$ (Prognostic factors)	$Z - score$
$1-10$	0.7964 ± 0.098	0.9952 ± 0.0112 (A, T, N, G, ER, M, Tr)	$2.0054(p < 0.05)$
$11-20$	0.7309 ± 0.06	0.8908 ± 0.0451 (A, T, N, G, M)	$2.3085(p < 0.01)$
$21-30$	0.7305 ± 0.053	0.8051 ± 0.0470 (A, T, N, G, Tr)	$1.0568(p = 0.14)$
$31-40$	0.7382 ± 0.047	0.7884 ± 0.0459 (A, T, N, G, Tr)	$0.7608(p = 0.22)$
$41-50$	0.7331 ± 0.045	0.7663 ± 0.0446 (A, T, N, G, Tr)	$0.5240(p = 0.30)$
$51-60$	0.7263 ± 0.043	0.7661 ± 0.0411 (A, T, N, G, H)	$0.6659(p = 0.25)$
$61-\text{MFT}^4$	0.6762 ± 0.042	0.8842 ± 0.0265 (A, T, N, G)	$4.1884(p < 0.0001)$
$21-60$	0.7316 ± 0.048	0.7812 ± 0.022	$1.5379(p = 0.06)$
$1-\text{MFT}^4$	0.7669 ± 0.0178	0.8497 ± 0.0153	$3.5276(p < 0.001)$

Table 3: The results obtained for the predictions of survival by time indicated as areas under the ROC curves (AUC) for neural networks (ANN) and for the Cox regression model. In the third column the prognostic factors used for the different intervals for the case of the neural network are also shown. The maximum follow-up time is indicated as MFT. In the last columns the Z-score and the corresponding significance value (p-value) for the difference between the two models are shown

to maximum follow-up time), the predictive accuracy was significantly better than the one obtained by using Cox regression model ($p - value < 0.001$). Five prognostic factors were found to be significative and commons for both neural network and Cox regression models (A, T, N, G, Tr). Besides, the Cox regression procedure selected the prognostic factor H. On the other hand, the time dependent analysis shows (Table 3) that the neural network outperforms significantly the Cox model mainly for short time intervals $1-10$ and $11-20$ months. For the periods between $21-30$, $31-40$, $41-50$, and $51-60$ the difference between NN and Cox was not statistically significant and this might be because a reduced number of cases is included in each of these intervals. When the data was grouped together between the interval $21-60$ the difference achieves nearly statistical significance ($p - value = 0.06$). In the last time interval (61 - MFT) the neural networks also outperforms the Cox model (difference between the AUC $= 0.109$; $p - value < 0.0001$). Regarding the prognostic factors considered optimal for the prediction of the relapse using the neural architectures, it is worthwhile the inclusion of menopausal status (M) for the early intervals $1-10$ and $1-20$ months and that in all the cases the optimal neural networks did not use H as a prognosis covariate.

3 Conclusion

Prognostic factors in breast cancer provide information to patients about the recurrence likelihood of the disease and, more important, assist the clinicians in the selection of appropriate adjuvant treatments for the individual patients. From a biological point of view, the identification of good prognostic factors supplies information about the natural history of the disease. Cox multivariate analysis has been accepted as the gold-standard in methods of prognostic factors identification. However, Cox multivariate analysis involves that some assumptions need to be made (e.g. the relative risk between the hazard rate for two subjects are constant over time), what makes the Cox application to biological systems mostly inadequate. In this study, using a classical set of prognostic factors an approach based on artificial neural networks provides better survival predictions than those obtained by applying the Cox multivariate analysis. This better prognosis accuracy is specially relevant in the first time-interval 1 - 10 months, with 0.99 for AUC value, in the second interval under study 11 - 20 months where AUC is 0.89 , for the last interval considered, 61 - MFT, where the AUC is 0.88. This study continues to confirm the advantage of using NNs models over the standard Cox regression tool for breast cancer prognosis.

References

[1] E. Biganzoli, P. Boracchi, D. Coradini, M. Daidone, and E. Marubini. Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors. *Ann Oncol*, 14:1484–1493, 2003.

[2] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini. Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach. *Stat Med*, 17:1169–1186, 1998.

[3] P. Boracchi, E. Biganzoli, and E. Marubini. Modelling cause-specific hazards with radial basis function artificial neural networks: application to 2233 breast cancer patients. *Stat Med*, 20:3677–3694, 2001.

[4] L. Bottaci, P. Drew, J. Hartley, M. Hadfield, R. Farouk, P. Lee, I. Macintyre, G. Duthie, and J. Monson. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*, 350:469–472, 1997.

[5] S. Brown, A. Branford, and W. Moran. On the use of artificial neural networks for the analysis of survival data. *IEEE Trans Neu Net*, 8:1071 –1077, 1997.

[6] H. Burke, P. Goodman, D. Rosen, D. Henson, J. Weinstein, F. Harrel, J. Marks, D. Winchester, and D. Bostwick. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer*, 79, 1997.

[7] M. Colleoni, N. Rotmensz, C. Robertson, L. Orlando, G. Viale, G. Renne, A. Luini, P. Veronesi, M. Intra, R. Orecchia, G. Catalano, V. Galimberti, F. Nole, G. Martinelli, and A. Goldhirsch. Very young women (¡35 years) with operable breast cancer: features of disease at presentation. *Ann Oncol*, 13:273–279, 2002.

[8] D. Cox. Regression models and life tables. *J R Stat Soc*, 34:187–202, 1972.

[9] M. De Laurentis, S. De Placido, A. Bianco, G. Clark, and P. Ravdin. A prognostic model that makes quantitative estimates of probability of relapse for breast cancer patients. *Clin Cancer Res*, 5:4133–4139, 1999.

[10] M. De Laurentis and P. Ravdin. A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Lett*, 77:127–138, 1994.

[11] S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inf*, 35:352–359, 2002.

[12] D. Faraggi, R. Simon, E. Yaskil, and A. Kramar. Bayesian neural network models for censored data. *Biometrica J.*, 5:519–532, 1997.

[13] P. Fitzgibbons, D. Page, D. Weaver, A. Thor, D. Allred, G. Clark, S. Ruby, F. O'Malley, J. Simpson, J. Connolly, D. Hayes, S. Edge, A. Lichter, and S. Schnitt. Prognostic factors in breast cancer. college of american pathologists consensus statement 1999. *Arch Pathol Lab Med*, 124:966, 2000.

[14] J. Fox. *An R and S-Plus companion to applied regression.* Sage Pub., 2002.

[15] R. Golden. *Mathematical Methods for Neural Network Analysis and Design.* MIT Press, 1996.

[16] M.K. Gospodarowicz, D.E. Henson, V.P. Hutter, B. O'Sullivan, L.H. Sobin, and Ch. Wittekind. *Prognostic Factors in Cancer, 2nd Edition.* John Wiley & Sons, Inc., 2001.

[17] M. Greiner, D. Pfeiffer, and R. Smith. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*, 45:23–41, 2000.

[18] J. Hanley and B. McNeil. A method for comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148:839–843, 1983.

[19] F. Harrell, R. Califf, D. Pryor, K. Lee, and R. Rosati. Evaluating the yield of medical tests. *J Am Med Assoc*, 247:2543–2546, 1982.

[20] F. Harrell, K. Lee, R. Califf, D. Pryor, and R. Rosati. Regression modeling strategies for improved prognostic prediction. *Stat Med*, 3:143–152, 1984.

[21] S. Haykin. *Neural networks: A Comprehensive Foundation.* Mc Millan, 1994.

[22] J. Jerez, J. Gómez, G. Ramos, J. Muñoz, and E. Alba. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif Intell Med*, 27:45–63, 2003.

[23] N. Kroman, M. Jensen, J. Wohlfahrt, H. Mouridsen, P. Andersen, and M. Melbye. Factors influencing the effect of age on prognosis in breast cancer: population based study. *BMJ*, 320:474–478, 2000.

[24] K. Liestol and P. Andersen. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Stat Med*, 21:3701–3714, 2002.

[25] M. Lundin, J. Lundin, H. Burke, S. Toikkanen, Pylkkänen, and H. Joensuu. Artificial neural networks applied to survival prediction in breast cancer. *Oncology*, 57:281–286, 1999.

[26] L. Ohno-Machado. A comparison of cox proportional hazards and artificial neural network models for medical prognosis. *Comput Biol Med*, 27:55–65, 1997.

[27] P. Ravdin and G. Clark. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat*, 22:285–293, 1992.

[28] R. Ripley, A. Harris, and L. Tarassenko. Neural network models for breast cancer prognosis. *Neural Comput. Appl.*, 7:367–375, 1998.

[29] J. Shao and D. Tu. *The Jacknife and Bootstrap.* Springer-Verlag, 1995.

[30] T. Therneau and P. Grambsch. *Modelling survival data: Extending survival data.* Springer-Verlag, 2000.

[31] A. Xiang, P. Lapuerta, A. Ryutov, J. Buckley, and S. Azen. Comparison of the performance of neural network methods and cox regression for censored survival data. *Comput Stat and Data Anal*, 34:243–257, 2000.