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

An Artificial Neural Network for Analysing the Survival of Patients with Colorectal Cancer

R Bittern[1], A Cuschieri[1], S D Dolgobrodov[2], R Marshall[2], P Moore[2], and R J C Steele[1]

1 Department of Surgery and Molecular Oncology
The University of Dundee, Dundee DD1 9SY, Scotland
2 School of Physics and Astronomy
The University of Manchester, Manchester M13 9PL, England

Abstract. An internet/web based artificial neural network has been developed for use by practicing clinical oncologists and medical researchers as part of a programme to aid decision making and eventually, the management and treatment of individual patients with colorectal cancer. We have configured and implemented a Partial Likelihood Artificial Neural Network (PLANN) and trained it to predict cancer related survival in patients with confirmed colorectal cancer using a database provided by the Clinical Resource and Audit Group (CRAG) in Scotland. The reliability of the trained PLANN was evaluated against Kaplan-Meier (KM) actual survival plots and shows close agreement with them.

1 Introduction

We have applied artificial neural networks (ANNs) and their associated analytical techniques to healthcare, with special reference to patients suffering from common solid cancers. There is increasing complexity in the staging and management of these cancers, requiring specialist, multidisciplinary knowledge, and management. We believe that analytical systems such as these will become more readily available to clinicians with the emergence of web and grid-secure technology, which has the potential to link large clinical and scientific data sets of cancer patients from various sources and institutions.

To date, ANNs of varying complexity and types have been used, mainly in clinical research rather than routine clinical oncology. Their usefulness has been investigated in the diagnosis, spread of the disease and prognosis in breast, ovarian, gastrointestinal, bronchial, prostatic and ovarian cancers [1-3]. In breast and colorectal cancers, ANNs have been shown to be significantly more accurate in predicting survival than in predicting spread from the primary cancer site [4]. To date, there have been no reported studies on the use of ANNs to formulate management plans for the treatment of patients with cancers, and this remains a long-term aim of the current interdisciplinary work by our group of oncologists in Dundee and physicists in Manchester. So far, we have trained the ANN by exposing it to sets of existing data on one type of solid cancer (colorectal), where the clinical outcome of the patients included in the data base is known over a 5-year follow-up period. This paper deals with the verification of the prediction of survival by our web-based system.

2 Patient Database and Methods

2.1 The data

Following the approval of the Clinical Resource and Audit Group (CRAG), the Scottish Colorectal Cancer Audit Database was used for this analysis. The significant event of interest in the study was defined as a cancer-related death occurring within 5 years of clinical follow-up beginning from the date of first diagnosis. Patients who fell into this category were designated as non-censored, they rest being censored. Sixteen categorical variables were selected and used for the ANN training. These are shown in Table 1. The 'Age group' variable had 6 attributes (< 50, 50–59, 60–69, 70–79, 80–89, 90 or older). In parts, the 16 variable data set was incomplete and so we kept only those records with known parameters for the most important variables as judged by the clinicians involved in the study. This led to a sub-set of about 1500 patient records for analysis. These parameters were age, Duke's stage, number of positive nodes, vascular pedicle node identified, chest x-ray, liver US/CT scan, laparoscopy. For the rest of the set, any missing values were set to be the mean of the particular variable.

N	Variable	Attributes
1	Age group	6
2	Duke's stage	A, B, C, D
3	No. of +ve lymph nodes	1-10
4	Vascular pedicel node identified	2
5	Chest x-ray	2
6	Liver US seen (abdomen)	2
7	Laparoscopy	2
8	Operation intent	3
9	Weight loss	3
10	Radiotherapy	3
11	Chemotherapy	3
12	Tumour size in bands	3
13	Tumour differentiation	2
14	Site group	4
15	Anastomosis leak	2
16	Clinical trial	2

Table 1: Variables recorded in the colorectal database

2.2 Modeling survival

Survival modeling was based on a Partial Likelihood Artificial Neural Network (PLANN) [5, 9], which attempts to solve problems associated with censoring and classification. The base element of the ANN is used to process the data incorporated in a Multilayer Perceptron (MLP) with a sigmoid hidden function. The actual architecture used is a variant of this, known as 'cascade multilayer architecture'. A single output unit computes the conditional failure probability. The input layer has

units for time and the covariates plus one bias unit. One of the advantages of such a model is the ease of incorporation of time-dependent covariates, since each subject is represented, for each interval, by one input sector which can change across intervals. The choice of this network configuration ensured maximal predictive capability.

The probability of dying (hazard) for each patient is computed by having an additional input to the neural network specifying the time interval required, and entering each patient into the training set for each time interval until death or the end of follow-up. This method has only a small bias due to censoring and it allows usage of covariates that are dynamically changing with time. It has been used by Biganzoli et al [5], Lisboa et al [9] and is reviewed by Ripley and Ripley [10]. It is also known as the 'chain-binomial' model or Partial Logistic Artificial Neural Network (PLANN). This ANN has one output – a binary target with possible values 1 or 0 and estimates a discrete time hazard for the particular time interval.

The survival probability was calculated from the estimated discrete time hazards by multiplying it by the conditionals for survival over successive time intervals, each one treated as independent. This scheme classifies patients within every time period into either 'alive' or 'dead' provided that the data in the training set is well balanced and the distribution of both classes is uniform. However, the optimum network solution has a tendency to ignore the least represented class at a specific time intervals (e.g., class 'dead' at the beginning of the follow-up history and class 'alive' at the end of follow-up). This results in certain biases in the final classification, which have to be addressed by weighting the distribution from the model by the log-likelihood [6, 7].

The final estimation of the ANN output is corrected using the Bayes Theorem as in [8]. This equalisation procedure is crucial for modeling the survival because the network's estimates of the hazard in each time interval are multiplied successively to obtain the survival curve. Since patients with cancer are seen regularly, monthly in the first year and then quarterly if they appear to be progressing well, there are 28 time intervals per patient during a 5-year follow-up period, the hazards per time interval is very low especially in the first year causing small deviations which would otherwise result in an unacceptable bias in the survival estimates.

The process of choosing the optimal set of variables was based on (a) the opinion of clinicians and (b) a 5-fold cross-validation procedure [15]. As a result a total of 1558 records of patients referred between 1993 and 1998 with a follow-up of 60 months were chosen. The PLANN was trained using the combination of simple regularization and the 'early stopping' technique [15]. The same cross-validation technique was used to choose the number of hidden nodes (= 23).

Another important issue is the confidence assessment of the predicted survival curves. There are two common approaches used for the assignment of confidence intervals for feed-forward neural networks, namely 'delta' and Bayesian statistics [18, 19]. Another useful approach is the bootstrap estimation of maximum likelihood frame-work, which is easier to implement in terms of both stability of the algorithm and speed of convergence. The bootstrap method appears to provide more accurate confidence intervals then the delta method [16, 18] in addition bootstrapping improves stability. The bootstrapping methodology was implemented by training 200 separate samples of ANNs on a randomly sampled subsets from the whole set of

records for each ANN and averaging the resulting survival curves in every point of the follow-up.

Finally, the model was calibrated against real survival by the Kaplan-Meier (KM) method [17] by producing comparative survival KM Curves.

3 Results

3.1 Prediction of colorectal cancer-related survival

The modeling was assessed by considering four test groups of patients with known follow-up outcome and at different risk level of death from cancer:

A. Duke stage A, node negative, no evident metastatic disease and considered to have had a definitively curative operation
B. Duke stage B, node negative, no evident metastatic disease and considered to have had a potentially curative operation
C. Dukes stage C but without evident metastatic disease, considered to have complete resection (no residual tumour) but a potentially non-curative operation
D. Duke stage D or extensive nodal disease or hepatic/ peritoneal disease, considered to have had definitely non-curative interventions or no surgical intervention

Obviously the hazard increases moving down the groups A to D.

Each of four groups was divided into two age groups: younger and older than 80, giving 8 groups in total. The graphical presentations of this comparison between predicted (PLANN) and actual survival (KM) are shown in Figure 1, which shows the survival KM histograms for the lowest two risk groups (Groups A and B) in patients with resected Duke's A colorectal cancer, with a known outcome in two age groups: >80 years (grey histograms, lower) and <80 years (black histograms, upper). The corresponding ANN predictions for the probability of survival are shown as grey and black curves with standard error bars (using the Greenwood formula [18]). The corresponding results for groups C and D (not shown) show a monotonic reduced survival. The predictions are in very good agreement with actual survival in all the four cohorts and in each age group. The noticeable error for the group of 'old' patients in Group A can be explained by the paucity of cases recorded.

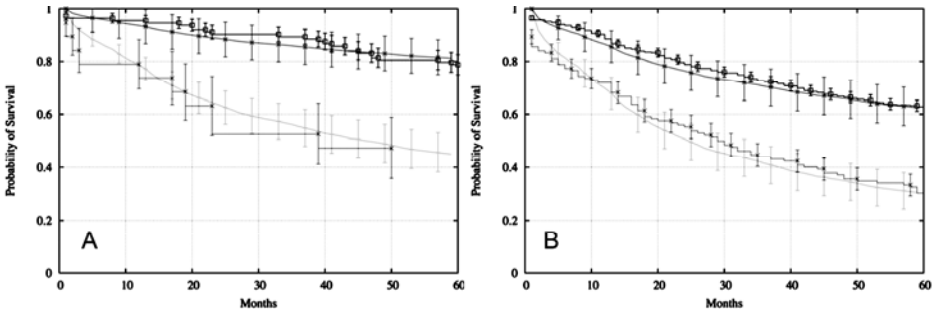

Fig. 1: Predicted survival curves for groups A and B compared with Kaplan-Meier

3.2 **Safe computer access to the system**

We have also designed a secure web based interface between the ANN and clinicians with kernel software, including higher sophisticated grid technology for data collection from remote data bases, training or re-training the PLANN using distributed computer resources and finally, provision of web access interface for *bona-fide* users. The clinicians will eventually be able to compare the curves for different management profiles. The web part of the system has been written in Perl cgi scripts and uses GRIDSITE server technology.

4 Discussion and conclusions

The PLANN model that has been implemented works well in most cases – as shown by the good agreement with the actual survival of the patient cohort used in the study. In addition, we can account for instances where the output predictions obtained were not in perfect agreement with actual survival of the patients. The root cause of these problems is the inconsistency and incompleteness of the data used to train the PLANN, resulting in unduly large variance. Regrettably this problem is inherent to all existing clinical databases collected centrally over many years within the NHS system. If ANNs are to be used more extensively in the future as is being predicted by the emergence of the GRID technology and associated middleware, this problem has to be resolved and the quality of data collection entry and management improved significantly beyond the existing level. The present situation undoubtedly inhibits progress in ANN modeling and development and possibly explains why these and other mathematical models are not widely used in practice by clinicians [12]. They have a reputation for unreliability and are regarded with suspicion the clinicians. This situation can be reversed by further study and improved data recording.

The present stage in the PLANN project can be regarded as a base platform for a system that could reliably stage, prognosticate and eventually produce an algorithm for the management of individual patients with cancer. We have produced a reliable and practical (easy to use) web interface that is already being used by clinicians. As the system is refined, it should be able to meet the requirements for the other objective – the optimal management of cancer patients commensurate with the stage of their disease. We believe that ANNs may also be developed and used for a variety of chronic disorders, e.g., cardiovascular disease, diabetes etc. with the benefit of standardisation that such systems bring to the treatment and management of patients.

We think that the clinical research potential of ANNs has been largely overlooked. It is not inconceivable that these systems as they become more advanced and hybrid (to include fuzzy logic) may eventually be developed to a stage that would enable clinical researchers to deploy them in 'hypothetical virtual human experiments' with various parameters applied to particular cases to predict better results of treatment or compare therapeutic options. This is probably some time off the present, but before then ANNs should be used in the analysis of on-going and completed randomised clinical trials as an additional means of data analysis to the current multivariate linear regression analyses systems.

References

[1] R. Dybkowski, Neural Computation in Medicine: Perspectives and Prospectives. In H. Malgren, M Borga, L Niklasson, editors, *proceedings of the ANNIMAB-1 Conference (Artificial Neural Networks in Medicine and Biology)*, Goteborg, 13-6 May 2000, Springer Verlag, pp. 26-36, 2000.

[2] P. J. G. Lisboa, A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks,* 15 (1): 9-37, 2002.

[3] W. G. Baxt, Application of artificial neural networks to clinical medicine. *Lancet,* 346:1135-38, 1995.

[4] R. N. Naguib and G. V. Sherb et al. Artificial neural networks in cancer research. *Pathobiology*, 65(3):129-139, 1997

[5] E. Biganzoli, P. Boracchi, L. Mariani, et al. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Stat Med,* 17:1169-86, 1998.

[6] D. Lowe and A. R. Webb, Exploiting prior knowledge in network optimization: an illustration from medical prognosis. *Network* ,1:299-323, 1990.

[7] P. J. G. Lisboa, A Vellido and H. Wong, Bias reduction in skewed binary classification with Bayesian neural networks, *Neural Networks,* 13:407-410, 2000.

[8] L. Tarassenko, A guide to neural computing applications, London, Willey, 1998.

[9] P. J. G. Lisboa, H. Wong, P. Harris, R. Swindell, A Bayesian neural network approach for modeling censored data with an application to prognosis after surgery for breast cancer. *Artificial Intelligence in Medicine,* 28:1-25, 2003.

[10] B. D. Ripley and R. M. Ripley, Neural networks as statistical methods in survival analysis. In R. Dybowski and V. Gant, editors, Artificial Neural Networks: *Prospects for Medicine*, Landes Biosciences Publishers, 1998.

[11] M. Mϕller, A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525-33, 1993.

[12] G. Schwarzer, W. Vach, W. M. Schumacher, On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. Tech. rep., Center for Data Analysis and Model Building, University of Freiburg, 1997.

[13] L. Breiman, Bagging Predictors, *Machine Learning*, 24:123-140, 1996.

[14] B. Cheng and D. M. Titterington, Neural networks: A review from a statistical perspective. *Statistical Science*, 9:2-54, 1994.

[15] C. M. Bishop, *Neural network for pattern recognition*. Oxford: Clarendon Press; 1995.

[16] G. E. Heskes, Practical confidence and prediction intervals, in M. Mozer, M. Jordan & T. Petch, eds, *Advances in Neural Information Processing Systems,* 9, MIT Press, Cambridge, MA, pp. 176-182, 1997.

[17] D. Collet, *Modeling survival data in medical research.* Chapman & Hall, London, 1994.

[18] R. Dybowski, S. Roberts, Confidence intervals and prediction intervals for feed-forward neural networks. In R. Dybowski, V. Gant editors, *Clinical Applications of Artificial Neural Networks*. Cambridge: Cambridge University Press, pp. 298-326, 2001.

[19] G. Papadopoulos, P. J. Edwards, A. F. Murray, *Confidence estimation methods for neural networks: a practical comparison*, IEEE transactions on neural networks, 12(6):1278-87, 2001.