

Artificial Neural Networks in Cancer Management

R. Bittern¹, A. Cuschieri¹, S. G. Dolgobrodov², R. Marshall², P. Moore².

¹) Department of Surgery, Ninewells The University of Dundee,

²)The University of Manchester

Abstract

An Instrument for Prediction of Cancer-Related Survival (CIPCS) has been implemented. The intent of this system is to provide an Internet/web access for practical clinicians to assist in making prognosis/decision in cancer patient treatment or management.

1 Introduction

The project was approved and funded in 2002 as part the MRC and PPARC e-Science programme. The stated aims of the project are to apply neural network techniques to the various databases that were held by or were available to the clinicians in Dundee.

We will discuss problems in the databases themselves and also in the procedures for getting the data to the analysts. Software code is now in action and the results are emerging from the studies. The first results of these analyses will be presented here. The response of the Manchester scientists to the data flow and the response of the clinicians to the first results have had an impact on the way the project is being handled and also on future plans. The original project, as approved, was based on e-Science and had no Grid content. But the Manchester part of the project team works in close proximity with local Grid experts who are developing Grid projects for particle physics. The assimilation of some of these Grid procedures into this project is now almost inevitable during the lifetime of this project.

The initial project team of four (two co-PIs and two PDRAs) was augmented by the award of a PPARC e-Science studentship to work on the project, starting in

Autumn 2002. The next stage of planned growth is the formation of a larger consortium bringing in a wider range of skills and resources.

2 Methodology and data exploitation

2.1 Clinical aspects

The aim of this research is to apply artificial neural networks (ANNs) and their associated analysis techniques to Health care, specifically to the management of breast and colorectal cancer patients. This particular problem has been posed by the clinical collaborators and represents a classic case where the solution to the medical problem appears to lie in the mathematics and information technology of the sort used by physicists. Grid secure technology can be used to link various sources of information (databases from the Information and Statistics Division, Edinburgh and Ninewells Hospital, Dundee) about cancer patients to ANN computer frame and vice versa. The ANN has been trained by exposing it to sets of existing data where the outcome (the follow up history of cancer patient) is known.

To date, ANNs of varying complexity and types have been used in research diagnosis and assessing the spread of the cancer. They have also been used in the overall prognosis of breast, ovarian, gastrointestinal, bronchial, prostate and ovarian cancers [1], [2], [3]. In breast and colorectal cancers, ANNs have been shown to be significantly more accurate in predicting survival than gauging the degree of advancement of the disease by measuring the spread of the cancer away from the

principal site [4]. There have been no studies on use of ANNs to formulate a management plan for the treatment of patients with cancers. This is of seminal importance in view of the multi-modality treatment required for cancers. The development of such a system is highly sought after by practical clinicians and is technically possible to realise in practice. This is a prime objective of the study.

It has already been shown [2] that the results obtained from ANNs from one institution can be applied to other centres. So, the project should lead to a better standardisation of treatment for these types of cancer. Patients will benefit by getting the right treatment combination to maximise duration and quality of life, and those that are incurable with limited expected survival, would be spared the morbidity of unnecessary interventions. As a consequence of this, the available resource for the treatment of patients within the NHS budget could be prioritised on a rational basis.

2.2 Data description

The data obtained from Scottish Colorectal cancer audit database (CRAG) were used for this study. The event of interest is defined as a cancer related death within 5 years of observation. Those patients satisfying this condition are non-censored otherwise they are treated as censored. Table 1 shows the 18 categorical variables collected for the ANN training. The first seven variables, (“Age”, “Duke’s stage”, “Number of nodes positive”, “Vascular pedicel node identified”, “Chest x-ray result”, “Liver US scan of abdomen”, “Laparoscopy”) are assumed as the most important according to clinical experience.

Table 1
Variables recorded in the colorectal DB

Variable	Attributes
Age	0–100
Duke’s stage	A,B,C,D
Number of nodes positive	1–10
Vascular pedicel node identified	2
Chest x-ray	2
Liver US scan of abdomen	2
Laparoscopy	2
Operation intent	3
Weight loss	3
Radiotherapy	3
Chemotherapy	3
Tumour size in bands	3
Tumour differentiation	2
Site group	4
Anastomosis leak	2
Clinical trial	2

Because the original data set contains missing or unknown records, it was decided to keep only those records with known values in the above group of most important variables. For the remaining variables any missing parameters were set to the average value of that particular variable.

Regular discussions between clinicians and scientists have been essential in helping the project team to discuss which parameters are the most important in creating the ANN. Prognostic variables have been ranked and weighted by the clinicians in Dundee.

2.3 Modelling survival

Survival modelling is based on Partial Likelihood Artificial Neural Network (PLANN) [5], [6]. The ANN model attempts to solve problems with both censoring and classification. As the base element of ANN to process the data involved an Multilayer Perceptron (MLP) with sigmoidal hidden node function is used. The actual architecture is a variant of this and is called “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 model is the ease of incorporating the time-dependant covariates, since each subject is represented, for each interval, by one input vector, which can change across intervals. The choice of the best network configuration was finally based on maximisation of its 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 t_k which is required, and entering each patient into the training set for each time interval until death or until the end of follow-up. This method has only a small bias due to censoring, and it also allows covariates that are dynamically changing within time to be used. It has been used by Biganzoli et al. [5]; see also Lisboa et al. [6] and the review by Ripley [7]. It is sometimes known as a ‘‘chain-binomial’’ model or Partial Logistic Artificial Neural Network (PLANN) model. The ANN has one output, binary targets and estimates the discrete time hazard h which for the standard MLP in analytically equations is

$$\frac{h_p(x_p, t_k)}{1 - h_p(x_p, t_k)} = \exp \left(\sum_{h=1}^{N_h} w_{hg} \left(\sum_{i=1}^{N_i} w_{ih} x_{pi} + wt_k + b_h \right) + b \right), \quad (1)$$

where the indices i and h denote the input and hidden node layers, respectively, p denotes the individual patient record, x_p is a static covariate vector containing a set of explanatory variables extracted from the patient record, t_k is the k -th time interval, and the activation function $g(a) = 1/(1 + e^{-a})$. The survivorship S is calculated from the estimated discrete time hazard by multiplying the conditionals for survival over successive time intervals, treated as independent:

$$S(t_k) = \prod_{l=1}^k P(t > t_l | t > t_{l-1}) = \prod_{l=1}^k (1 - h(t_l)). \quad (2)$$

Thus, a patient dying in the i -th interval contributes $\log[h(t_i)(1 - h(t_{i-1})) \dots (1 - h(t_1))]$ to the log-likelihood, and a patient lost to follow up in that interval $\log[(1 - h(t_{i-1})) \dots (1 - h(t_1))]$, and from this the negative log-likelihood G can be computed. We can

expand the contribution to the log-likelihood as a sum of or over the intervals for which that patient is at risk and over all N patients under consideration:

$$G = - \sum_{p=1}^N \sum_{k=1}^{t_i} [d_{pk} \log(h_p(x_p, t_k)) + (1 - d_{pk}) \log(1 - h_p(x_p, t_k))], \quad (3)$$

where t_i is the time period when the patient was last observed, and $d_{pk} = 1$ if death attributed to cancer in time interval $t_{k-1} < t \leq t_k$ is observed and $d_{pk} = 0$ otherwise. The regularisation term $\sum_{i,j} \lambda_i w_{ij}^2$ has been added to (3) which gives the cost function $\bar{E} = G + \sum_{i,j} \lambda_i w_{ij}^2$. The network weights w_{ij} are estimated by minimising this function by means of scaled conjugate gradient algorithm [8].

Due to instability of predictions by ANNs, estimations of the generalisation performance of a single ANN should be based on a k -fold cross validation [12],[7]. The process of choosing the optimal set of variables from dozens in the original data base was based on a) opinion of clinicians, b) 5-fold cross-validation procedure [12]. As a result, a total of 1558 records of patients referred between 1993 and 1998 with a follow up history of 60 months have been chosen. The ANN was trained using the combination of simple regularization and the ‘‘early stopping’’ technique [12]. The same cross-validation technique has been used to choose the optimal value for the λ and the number of hidden nodes. The results were found to be $\lambda \approx 0.001$ and $N_h \approx 23$ respectively. Precautionary measures have been made to avoid triggering a premature end to training caused by local non-monotonic behaviour of the error on the validation set. The minimum of the cost function was searched over a wide range of values with the smallest one ever met being chosen.

Despite the existence of very sophisticated techniques ANN training process still remains a ‘‘state of the art’’ in many aspects. This assumes careful consideration of the results of training different ANNs with variable parameters of architecture, regularization and the set of covariates and selecting the best ones relatively to maximum likelihood or evidence when using a Bayesian framework.

One of the problems is the instability of the predic-

tion by the resulting system due to the variance of data used for the training process (see [10]) and the initial values in the weight matrix (usually randomly generated). This instability problem can be attacked by building an ensemble of ANNs and aggregating the results of the networks in the ensemble to produce reliable predictions.

Another important issue is the confidence assessment of the final survival curves. There are several approaches to the assignment of confidence intervals to feed-forward neural networks, namely, “delta” and Bayesian statistics [14], [15]. A popular approach is the bootstrap estimation in maximum likelihood framework (MLE). In contrast with the former approaches, MLE is easier to implement in terms of both stability of the algorithm and speed of convergence. The bootstrap method has been reported to provide more accurate confidence intervals than the delta method [14],[13]. In addition bootstrapping helps to deal with the problem of stability mentioned above.

The bootstrapping methodology has been implemented by training 200 samples of ANNs and averaging the resulting survival curves in every time point through the follow up history. This also provides a natural method for estimating the confidence intervals due to data variance and variance due to random initial weight matrix.

3 Results

3.1 Prediction of Colorectal Cancer-Related Survival

The graphical presentations examples of the results of survival predictions are shown in Figures 1 and 2.

Figure 1 shows the assessment for the survival for 7 colorectal cancer patients with known follow up history. The abscissa is the time in months after date of diagnoses and the ordinate represents the assessment of the probability that the patient will survive to this time. The text legend shows the real survival or censoring time. Those patients whose death time is beyond the study term are marked with “s” in the legend and the arrows on the curves.

The group with higher risk is well pronounced (see

4 lower curves in figure 1). To some extent the curves may be interpreted as a real measure of the deterioration of the patient’s health stage (quality of life). The curves shown are obtained from the test set. These curves match those of classic Kaplan-Meier estimations for Survivor Function in accordant groups of risk.

Similar result for breast cancer patients are shown in Figure 2.

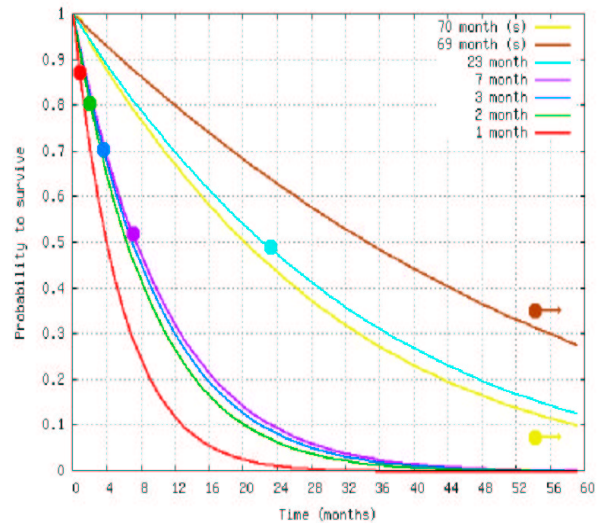


Figure 1: Prediction of the probability of survival for 7 colorectal cancer patients with known follow up history

3.2 Safe computer access to the system

The whole system has been aimed as a web based secure and safe accessible computer system. This includes higher sophisticated GRID technology for data collecting from remote data bases, training or retraining the ANN using distributed computer resources and finally providing web access interface for a users, clinical experts. Figure 3 shows an example screen shot of the interface. This includes the window with predictive survival curve in 60 months scaled time axis, folded input windows to choose the values from the predefined ones, and the frequency graphs for the parameters used. The web part of the system has been written in Perl

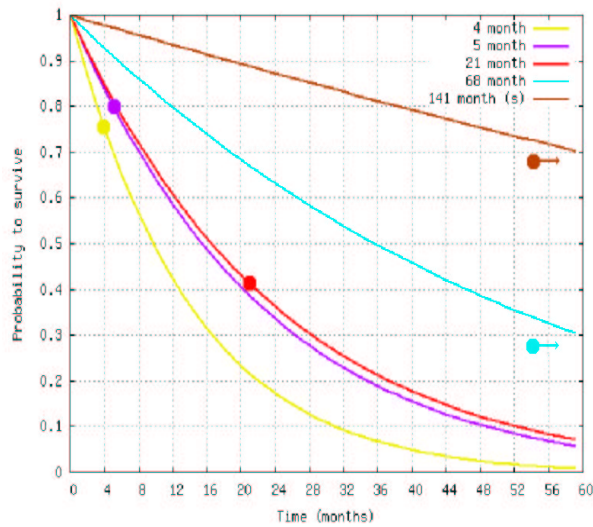


Figure 2: Prediction of the probability of survival for 5 breast cancer patients with known follow up history

and CGI scripts and uses GRIDSITE server technology. The full implementation of the GRID secure access and linkage with remote data bases yet have to be met.

4 Conclusion

The results obtained are not in total agreement with actual survival and sometimes they are not consistent with clinicians' expectations. There are many reasons for this. Despite the use of these sophisticated mathematical methods, a root cause of these problems is the inconsistency of the information in the data used to train the ANN. Also, the variables are sometimes misinterpreted when used for training. It is of critical importance to find a solution to these problems, even a partial one. At present one of the consequences is that ANNs and other intelligent and highly sophisticated tools are not widely used in practice by clinicians [9]. They have a reputation for unreliability and being somewhat of a "black box".

The novel methods introduced in this study are a) using the GRID for on-line access for the neural network processing of the medical data; b) implementation of

the use of vector decay shared by a group of covariates in automatic relevance determination and regularisation process during ANN training; c) implementation and use of time dependent covariates in ANNs.

The system currently under development can be treated as a prototype for the system which may serve as a practical Computer Instrument for Prediction of Cancer-Related Survival (CIPCS). The system also can help in improving primary trial and management of cancer patients (and also other types of diseases). Preliminary results shows that this system can be useful in daily routine practice of medical researchers and hospital clinicians.

Such a system can be also helpful for practical clinicians as instrument for "hypothetical experiments" with various parameters of clinical treatment applied to particular cases to predict better results of treatment and better management in every particular case.

Our experience shows that any similar attempt must be carried in a close collaboration with practical clinicians especially in selecting important variables and interpreting their impact. This allows the best results to be achieved, that are useful for practitioners and avoids misunderstandings between specialists in different science areas.

References

- [1] R. Dybowski, Neural Computation in Medicine: Perspectives and Prospects. In H. Malmgren, M. Borga, L. Niklasson (eds.) Proceedings of the ANNIMAB-1 Conference (Artificial Neural Networks in Medicine and Biology), Goteborg, 13-16 May 2000, Springer, 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] P. J. G. Lisboa, H. Wong, P. Harris, R. Swindell, A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artificial Intelligence in Medicine* 28:1–25, 2003.
- [7] 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.
- [8] M. Møller A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 6(4):525–33, 1993.
- [9] 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.
- [10] L. Breiman, Bagging Predictors, *Machine Learning* 24:123–140, 1996.
- [11] B. Cheng and D. M. Titterton, Neural networks: A review from a statistical perspective. *Statistical Science*, 9:2–54, 1994.
- [12] C. M. Bishop, *Neural network for pattern recognition*. Oxford: Clarendon Press; 1995.
- [13] G. E. Heskens, 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.
- [14] R. Dybowski, S. Roberts. Confidence intervals and prediction intervals for feed-forward neural networks. In R. Dybowski, V. Gant eds, *Clinical Applications of Artificial Neural Networks*. Cambridge: Cambridge University Press, pp 298–326, 2001.
- [15] G. Papadopoulos, P. J. Edwards, A. F. Murray, Confidence estimation methods for neural networks: a practical comparisons. *IEEE transactions on neural networks*, 12(6):1278–87, 2001.

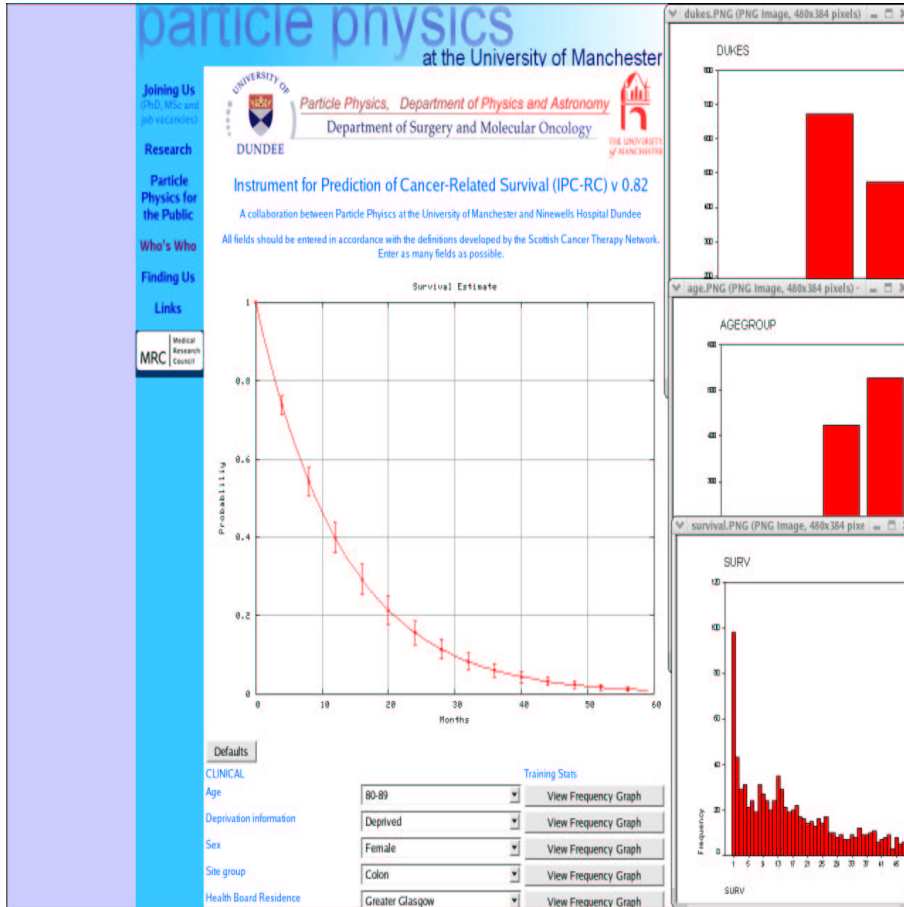


Figure 3: A screen shot of the user graphical web interface for the system access. This includes the window with predictive survival curve in 60 months scaled time axis, folded input windows to choose the values from the predefined ones, the and frequency graphs for the parameters used.