

Comparative study of intelligent models for the prediction of bladder cancer progression

M.F. ABBOD¹, D.A. LINKENS¹, J.W.F. CATTO² and F.C. HAMDY²

¹Department of Automatic Control and Systems Engineering, and

²The Academic Urology Unit, University of Sheffield, Sheffield, UK

Received September 6, 2005; Accepted October 10, 2005

Abstract. New techniques for the prediction of tumour behaviour are needed since statistical analysis has low accuracy and is not applicable to the individual. Artificial intelligence (AI) may provide suitable methods. We have compared the predictive accuracies of neuro-fuzzy modelling (NFM), artificial neural networks (ANN) and traditional statistical methods for the prediction of bladder cancer. Experimental molecular biomarkers, including p53 expression and gene methylation, and conventional clinicopathological data were studied in a cohort of 117 patients with bladder cancer. For all 3 methods, models were produced to predict the presence and timing of tumour progression. Both methods of AI predicted progression with an accuracy ranging from 88-100%, which was superior to logistic regression, and NFM appeared to be better than ANN at predicting the timing of progression.

Introduction

Transitional cell carcinoma (TCC) of the bladder is the fourth most common cancer amongst men in the UK (1). At presentation, 70% of TCC are superficial and non-invasive, which can be managed by local endoscopic resection and intra-vesical chemotherapy. Following treatment, these tumours require cystoscopic surveillance (2), as 50% will recur as similar non-invasive lesions, and a smaller percentage (20%) will progress to muscle invasion. Muscle invasive tumours have a poor prognosis (5-year survival rate of 50%) and require radical therapy if a cure is to be achieved (3). Following radical treatment, adjuvant chemotherapy can be used to reduce relapse and possibly mortality rates in selected patients with the highest risk of relapse (4).

The accurate prediction of future cancer behaviour would be of obvious benefit to both the patient and physician. Patients

with non-relapsing tumours could be safely reassured and discharged, while relapsing tumours could be treated more aggressively. The most reliable predictors of tumour behaviour are the pathological stage and grade at diagnosis (TNM classification) (5,6). Specific tumours also have additional prognostic information, including lymph node status in invasive disease and recurrence in superficial disease. While these parameters stratify patients into subgroups, it is impossible to predict individual tumour behaviour. The development of molecular medicine has yielded several new molecules that may be useful as predictive biomarkers. Some of the most biologically promising are the p53 gene expression and gene methylation. The p53 gene is mutated in over 50% of human cancers (6) and has been shown to predict recurrence and survival in bladder cancer (7). Gene methylation occurs in the majority of tumours and is associated with a poor outcome (8).

A solution to the problem of predicting tumour behaviour lies potentially within the interpretation of data. Traditional statistical methods, e.g. logistic regression (LR), produce probabilities of behaviour, which may be applicable to a population but are not predictive for an individual. Furthermore, their predictions are only accurate in 70% of tumours using the TNM classification (7). By using artificial intelligence (AI) methodologies, such as artificial neural networks (ANN) and neuro-fuzzy modelling (NFM), complex relationships between dependent and independent variables, in a population whose distribution may not be normal, can be identified. As a result, the prediction of biological behaviour from both physiological and pathological data can be performed.

ANNs, of which the most commonly used is the multi-layer perceptron (MLP), have been applied to clinical medicine since 1989 (9). However, ANNs are not without problems. They can be 'over-trained' to learn the inherent variation ('noise') of a sample population, and the network is hidden within a functional 'black box'. Thus, it is difficult to gain insight into the solution used to resolve the clinical data, making subsequent analysis (to ensure clinical sense prevails) and interrogation of new variables almost impossible. NFM is an alternate AI method, without many of the drawbacks of ANN.

Materials and methods

Patients. A total of 117 patients with primary TCC (chosen at random to represent the disease spectrum) were studied, of which 107 (92%) had tumours located in the bladder. Each

Correspondence to: Dr M.F. Abbod, Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK

E-mail: m.f.abbod@sheffield.ac.uk

Key words: artificial neural network, neuro-fuzzy modelling, bladder cancer, prognostic prediction, p53, mismatch repair, methylation

Table I. Input variables for the modelling methods.

Input variables		Scoring	
Analysis A			
Stage	Ta	T1	T2-4
Grade	1	2	3
Age	In years		
Sex	M/F		
Smoking exposure	Pack years		
Previous cancers (non-TCC)	0= None	1=1	2=>2
Analysis B-D			
p53	0 = Normal	1= Abnormal	
Methylation	percentage		
RARB	0= Abnormal	1= Normal	

Four analyses were performed (A-D). For analysis A, the inputs were the 6 conventional clinicopathological data. For analysis B, there was a p53 additional input, analysis C includes the conventional inputs plus % met, and analysis D is the same as C plus RARB.

tumour was excised at the Department of Urology, Royal Hallamshire Hospital, Sheffield, UK by either endoscopic resection (for pTa and pT1 disease), radical cystectomy (for muscle invasive disease) or nephro-ureterectomy (all upper tract TCC). Tumours were staged according to the 1997 TNM classification, and graded according to the 1973 WHO classification. Adjuvant intra-vesical chemo and BCG immunotherapy were used according to standard regimens (10). Following treatment, patients underwent surveillance for subsequent relapse according to risk for a median of 96 months (range 1-204 months). No patients with invasive TCC received adjuvant or neo-adjuvant systemic chemotherapy. The population studied was typical for bladder cancer with the majority of patients being male and

having a smoking history. In those patients with a smoking history, the cigarette exposure was calculated in pack years [number of years equivalent to 20 cigarettes per day; median 40 pack years (range, 0.5–112.0 pack years)]. Primary non-urothelial cancers had been diagnosed in 24 patients (20%), 6 of whom had more than one other malignancy. Progression was defined as the occurrence of a subsequent tumour of higher stage or grade to the initial TCC at diagnosis.

For each tumour, nuclear protein accumulation of p53 was established. Tumours with >30% positively staining nuclei were scored as abnormal by workers blinded to the clinical outcome of each tumour. Gene promoter methylation was investigated using methyl-specific PCR at 11 loci. The presence of methylation at each locus was regarded as abnormal, and an overall methylation index (MI) was calculated (no. methylated loci/no. successfully analysed loci) for each tumour. These methods and results are discussed in more detail elsewhere (9). For modelling, we investigated the MI of each tumour and the individual status of the RARB (retinoic acid receptor B gene) locus, which appeared to be a good prognostic marker using log-rank analysis.

Predictive models were developed for the progression of TCC of each patient. The input data shown in Table I were used in 5 separate combinations. Analysis A used only clinicopathological details, analysis B used p53 expression and clinicopathological data, analyses C and D used the MI of each tumour with and without RARB methylation (respectively) in addition to the data from A, while all of the molecular and clinicopathological data were used for analysis E. For each analysis, 2 different predictive models were developed. The first model was defined as a 'classifier' and predicted the risk of progression occurrence. The output was binary, and stratified patients into those with and without progression. The second model was defined as a 'predictor' and predicted the time of progression (in months) following initial resection. To compare the results of AI with traditional statistics, we used LR to similarly predict the time of tumour progression. The performances of each predictive method

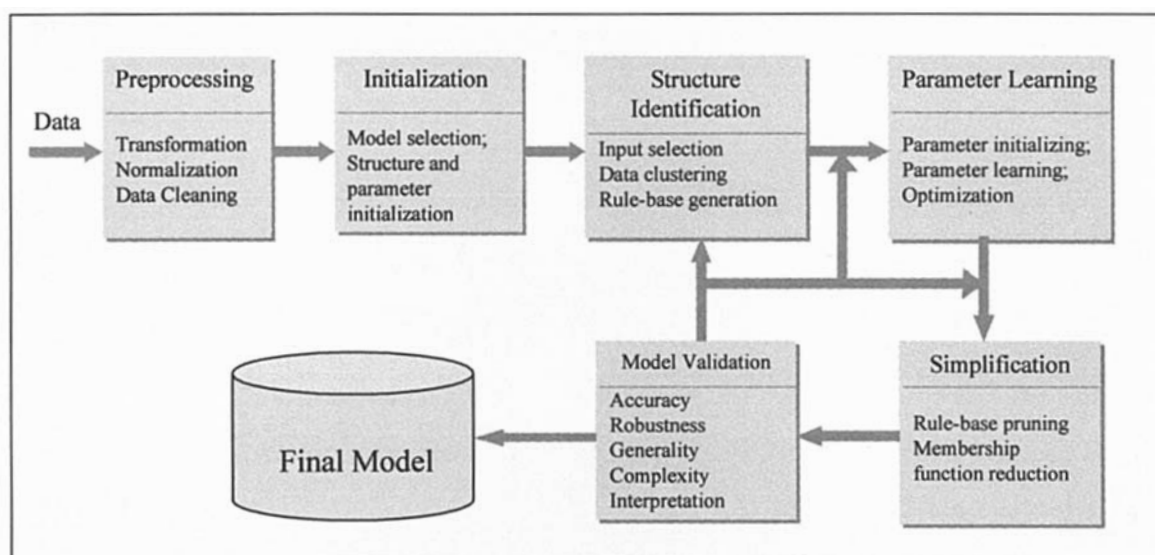


Figure 1. Schematic structure of the NFM suite of programs used to produce an optimal complexity model from the available data.

	ANN			NFM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
A	81	95	89	88	99	94
B	87	97	93	100	97	98
C	83	100	92	100	100	100
D	80	100	90	100	100	100

Table III. Time to tumour progression: results of the AI predictor models and LR.

Case	ANN				NFM				LR
	Training	Validation	Testing	Total	Training	Validation	Testing	Total	Total
A	10.05	15.71	25.95	9.01	5.08	19.63	14.12	5.31	13.42
B	9.44	12.02	21.54	9.85	1.42	26.53	25.37	5.18	13.26
C	9.79	13.93	23.71	8.01	2.29	31.99	25.37	5.17	19.12
D	9.08	13.33	18.06	8.35	3.26	17.81	21.77	5.09	18.12

The times to tumour progression are shown as root mean squares (difference between predicted and actual time of relapse). For ANN and NFM, after training and testing, the data was then analysed for the overall best-fit model.

were compared using a two-tailed t-test. A p-value <0.01 was taken as statistically significant.

Neuro-fuzzy modelling. The NFM analyses were performed via an extensive in-house suite of software developed in Matlab (11). A schematic diagram of this extensive suite is given in Fig. 1. The modelling procedure involves a number of iterative loops subsequent to careful data preparation and initialisation of the starting model structure and parameters. These loops refine the model parameters, simplify its structure and component terms to the minimum complexity consistent with the model (i.e. parsimonious modelling), and validate the results.

Ensemble modelling. Averaging the output of several different models is called an ensemble of models, or simply an ensemble. The idea of averaging different models was developed by the neural network community at the beginning of the 90s. It was pointed out that the generalization error of the ensemble is lower than the mean of the generalization error of single ensemble members (12).

Cross validation. In order to estimate the generalization error and select models for the final ensemble, a cross-validation scheme is used for model training (13). The cross-validation is done in several training rounds on different training sets, which increases the diversity of the ensemble and leads to better generalization. Another advantage of this method is to get an unbiased estimator of the ensemble generalization error and, at the same time, train the ensemble on the whole dataset. This is useful in situations where only a few data points are available.

Results

The AI modelling was performed on commercially available Matlab software. The fuzzy logic predictions of NFM were developed using Matlab and described in more detail elsewhere (11). The ANN model was a multi-layered perceptron with 15 hidden neurones. In each ANN session, 10 models were developed and the best one was chosen ('best fit'). Training was performed for a maximum of 50 iterative loops, and stopped once overtraining occurred. For both ANN and NFM, the data were analysed using cross-validation. Thus, the model was trained on 90% of data and tested on the remaining 10%. This was repeated 10 times, until the model had been tested on each sample. The 'best fit' ANN model and trained NFM model were then retested on the data to obtain a 'final' outcome.

The 117 patients with TCC represented a typical UK population (median age, 70 years; 65% male and 62% smokers). Smoking was significantly related to more advanced disease compared to non-smoking.

The results of the classifier models generated by using ANN and NFM are shown in Table II. The accuracy for ANN and NFM was 93% and 98% for analysis B and 90% and 100% for analysis D, respectively. Results for the predictor models are shown in comparison with LR (Table III). In each case, the difference between the actual and predicted time of progression is shown as a root mean square value (RMS). In all categories, the AI models perform better than LR, and NFM is more accurate than ANN. As demonstrated, ANN and NFM are significantly superior to LR. When ANN and NFM are directly compared, NFM is significantly better than ANN at predicting tumour progression. The predictions

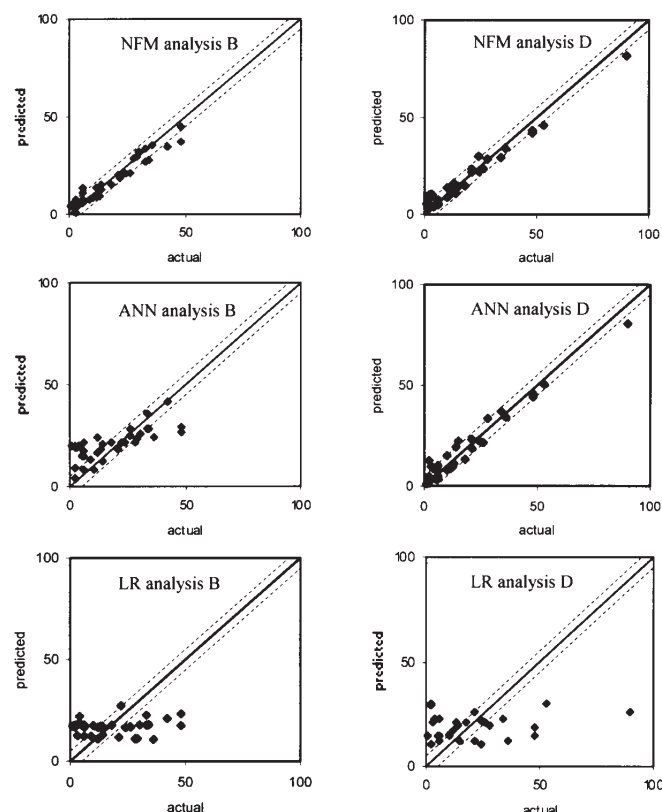


Figure 2. Scatter plots of actual and predicted progression times. In all six graphs each point represents the actual time of tumour progression (x axis; 0-100 months after surgery) against the predicted time of tumour progression (y axis; 0-100 months after surgery). A continuous line indicates the accuracy of each plot, and dashed lines indicate the $\pm 10\%$ range.

of all 3 methods are shown graphically as scatter plots in Fig. 2.

Discussion

The accurate prediction of an individual tumour response to treatment is a Holy Grail of oncology. Here, we have shown that NFM can predict tumour behaviour with greater accuracy than both ANN and LR. Until the advent of AI, the best method of predicting tumour behaviour was logistic regression. Using Fig. 2, if LR were applied in clinical practice, patients with late relapsing tumours (>40 months) would have had their most intensive cystoscopic surveillance too early for their actual relapse.

Previous authors have shown that ANN can predict tumour behaviour more accurately than LR and clinicians (7). We have again confirmed that ANN provides a powerful and accurate predictive method. However, unlike these previous

studies, we were able to compare ANN with NFM. NFM is a relatively novel modelling technique. A previous report and our current study have shown that NFM produces a significantly more accurate prediction than ANN and LR (Table III). In addition to accuracy, NFM has other benefits over ANN. Unlike the 'black-box' phenomena of ANN, the interpolation of rules in the NFM approach is transparent. By 'defuzzification,' qualitative modelling figures can be translated into understandable medical terms. Thus, NFM can incorporate expertise and allow predictions of outcome that result from changes in the value of individual inputs. These features make NFM an important tool with extensive clinical applications. While our predictions have been modelled using bladder cancer, these methods are transferable to many other human malignancies.

References

1. Quinn M: Cancer trends in England and Wales 1950-1999. In: SMPs. 66th edition. The Stationary Office, 2000.
2. Hall RR, Parmar MK, Richards AB and Smith PH: Proposal for changes in cystoscopic follow up of patients with bladder cancer and adjuvant intravesical chemotherapy. *Br Med J* 308: 257-260, 1994.
3. Stein JP, Lieskovsky G, Cote RJ, *et al*: Radical cystectomy in the treatment of invasive bladder cancer; long term results in 1054 patients. *J Clin Oncol* 19: 666-675, 2001.
4. Raghaven D, Quinn D, Skinner DG and Stein JP: Surgery and adjunctive chemotherapy for invasive bladder cancer. *Surg Oncol* 11: 55-63, 2002.
5. Sobin LH and Wittekind CH: TNM Classification of Malignant Tumours. 5th edition. John Wiley and Sons, New York, NY, 1997.
6. Parmar MKB, Freedman LS, Hargreave TB and Tolley DA: Prognostic factors for recurrence and follow up policies in the treatment of superficial bladder cancer: interim report from the British Medical Research Council subgroup on superficial bladder cancer (Urological cancer working party). *J Urol* 142: 284-288, 1989.
7. Burke HB, Goodman PH, Rosen DB, *et al*: Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 79: 857-862, 1997.
8. Catto JWF, Azzouzi A, Rehman I, *et al*: Promoter hypermethylation is associated with tumour location, stage and subsequent progression in transitional cell carcinoma. *J Oncol* (In press).
9. Cross SS, Harrison RF and Kennedy RL: Introduction to neural networks. *Lancet* 346: 1075-1079, 1995.
10. Catto JWF, Xinarianos G, Burton JL, *et al*: Differential expression of hMLH1 and hMSH2 is related to bladder cancer grade, stage and prognosis, but not microsatellite instability. *Int J Cancer* 105: 484, 2003.
11. Chen M and Linkens DA: A systematic neuro-fuzzy modelling framework with application to material property prediction. *IEEE Trans SMC, part B. Cybernetics* 31: 781-790, 2001.
12. Krogh A and Vedelsby J: Neural network ensembles, cross validation and active learning. In: *Advances in Neural Information Processing Systems 7*. MIT Press, 1995.
13. Hastie T, Tibshirani R and Friedman J: *The Elements of Statistical Learning*. Springer, 2001.