Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions

Leonardo Bottaci, Philip J Drew, John E Hartley, Matthew B Hadfield, Ridzuan Farouk, Peter W R Lee, Iain M C Macintyre, Graeme S Duthie, John R T Monson

Summary

Background Artificial neural networks are computer programs that can be used to discover complex relations within data sets. They permit the recognition of patterns in complex biological data sets that cannot be detected with conventional linear statistical analysis. One such complex problem is the prediction of outcome for individual patients treated for colorectal cancer. Predictions of outcome in such patients have traditionally been based on population statistics. However, these predictions have little meaning for the individual patient. We report the training of neural networks to predict outcome for individual patients from one institution and their predictive performance on data from a different institution in another region.

Methods 5-year follow-up data from 334 patients treated for colorectal cancer were used to train and validate six neural networks designed for the prediction of death within 9, 12, 15, 18, 21, and 24 months. The previously trained 12-month neural network was then applied to 2-year follow-up data from patients from a second institution; outcome was concealed. No further training of the neural network was undertaken. The network's predictions were compared with those of two consultant colorectal surgeons supplied with the same data.

Findings All six neural networks were able to achieve overall accuracy greater than 80% for the prediction of death for individual patients at institution I within 9, 12, 15, 18, 21, and 24 months. The mean sensitivity and specificity were 60% and 88%. When the neural network trained to predict death within 12 months was applied to data from the second institution, overall accuracy of 90% (95% CI 84–96) was achieved, compared with the overall accuracy of the colorectal surgeons of 79% (71–87) and 75% (66–84).

Interpretation The neural networks were able to predict outcome for individual patients with colorectal cancer much more accurately than the currently available clinicopathological methods. Once trained on data from one institution, the neural networks were able to predict outcome for patients from an unrelated institution.

Lancet 1997; **350:** 469–72

Department of Computer Science, University of Hull (L Bottaci PhD)**; University of Hull Academic Surgical Unit, Castle Hill Hospital, Hull HU16 5JQ** (P J Drew FRCSE, J E Hartley FRCS, M B Hadfield FRCS, R Farouk FRCSE, P W R Lee FRCS, G S Duthie FRCSE, Prof J R T Monson FRCS)**; and Western General Hospital, Edinburgh, UK** (I M C Macintyre FRCSE)

Correspondence to: Professor John R T Monson

Introduction

At present, it is extremely difficult to predict whether and when an individual patient will die after surgical treatment for colorectal cancer. The only available techniques rely on the calculation of the risk affecting large groups within the population as a whole. These methods have been used to provide predictions based on pathological staging techniques such as those of Dukes[1] and Jass and colleagues[2] and the identification of prognostic indicators by univariate and multivariate analysis.[3] For example, it is known that after treatment for a Dukes' C cancer the chance of surviving at least 5 years is about 50%.[1] However, the mean population survival is not a useful predictor of survival for the individual patient—many patients with Dukes' C cancers have survival that deviates widely from the mean. We applied artificial neural networks to the analysis of data from patients with colorectal cancer in an attempt to achieve accurate predictions of outcome for individual patients.

Methods

Development of a neural network

Design—We decided to use a fully connected multilayer feedforward network, since the analytical power of this type of network is good and is well understood.[4,5] The networks were constructed by means of general-purpose neural-network software[6] and executed on a Sun Workstation under UNIX. The number of units in the input layer was determined by the number of input data values. A single output unit was used. The number of units in the remaining middle layer was chosen by experimentation. Several different networks were constructed containing varying numbers (two to 15) of units in the middle layer. The logistic activation function was used with continuous output on the interval (0, 1).

Training—Data collected prospectively over 5 years on 334 patients treated for colorectal cancer were used to train and validate the networks. The selected variables (panel) were separated into categorical, ordinal, or ratio-scale values giving a total input of 42 clinicopathological variables. The learning rule used was back propagation of error, which adjusts the internal

Clinicopathological factors used to train and validate the artificial neural networks

Patient dependent	Clinician dependent
Age	Liver metastases on preoperative ultrasound scan
Sex	Liver metastases at operation
Abdominal pain	Operation type
Altered bowel habit	Surgeon rank
Weight loss	Tumour fixity
Obstruction	Involvement of adjacent organs
Palpable mass	Peritoneal involvement
Mucus per rectum	Tumour perforation
Bleeding per rectum	Resection judged curative
Anaemia	Anastomosis type
Family history	Blood transfusion
History of inflammatory bowel disease	Adjuvant therapy
	Dukes' stage

	Dukes' stage			
	A	B	C	D
Initial proportion	14%	46%	37%	3%
Mortality (% of stage group)				
9 months	7	8	28	44
12 months	7	11	32	44
15 months	7	13	39	44
18 months	9	13	42	44
21 months	9	16	47	67
24 months	11	18	51	67

Table 1: **Distribution by Dukes' stage and mortality by time and stage in institution I data set**

parameters of the network over the repeated training cycles to reduce the overall error.[5,7] Training was terminated when the sum of squares error with respect to the validation data set was at a minimum. We thus took care not to over-fit the training data.

Validation—Of the 334 patients in the training database, 50 were randomly selected for validation and the remaining 284 were used for training. To eliminate any possible bias, each network was trained and validated with 30 randomly selected training and validation data sets. The differences in network performance among these 30 selections were not significant.

Application—100 patients were then randomly selected from a database of prospective 2-year follow-up data from a second institution. Patients dying from causes other than colorectal cancer were excluded, which left a population of 92 patients. A neural network, trained and validated by means of the data from institution I to predict death within 12 months, was then used to predict outcome for this group. No retraining of the network was permitted. The data presented to the trained neural network were also presented to two consultant colorectal surgeons who were asked to predict which individual patients would die within 12 months.

To provide a valid comparison between the clinicians and the neural network and to give an indication of the clinical usefulness of the neural networks' predictions, results are expressed in terms of overall accuracy (sum of correct predictions of death or life divided by total number of predictions), sensitivity, specificity, positive and negative predictive values, and odds likelihood ratio with 95% CI where appropriate.[8] The likelihood ratio indicates the value of a test for increasing certainty about prediction or diagnosis. Effectively, the post-test odds are equivalent to the pretest odds multiplied by the likelihood ratio, which is calculated as sensitivity/(1−specificity). A McNemar test, a non-parametric test for two related dichotomous variables, was used to analyse the relative accuracy of the predictions from clinicians and neural networks. All statistical analysis was done on SPSS 6.1 for Windows.

Results

The distribution by Dukes' stage and relation to survival for the two data sets are given in tables 1 and 2. The neural networks were able to achieve an accuracy of prediction for mortality from colorectal cancer of more than 80% for all six time periods for patients from institution I (table 3). The probability that a network would correctly predict death within a given 3-month period (positive predictive value) varied from 61% to 71%, but the likelihood ratio remained high (table 3). In addition, the neural network was able to produce the highest likelihood ratio for the period (9 months) in which a prediction had the lowest probability of being correct (61%).

	Dukes' stage			
	A	B	C	Unknown
Initial number (% of total)	20 (22%)	35 (38%)	33 (36%)	4 (4%)
Number of deaths	0	0	5*	1†

*At 3, 4, 6, 9, and 11 months. †At 4 months.

Table 2: **Distribution by Dukes' stage and mortality in institution II data set**

The performance of the network previously trained on the data from institution I was significantly better than that of the two consultant surgeons when the data for institution II were presented for analysis (McNemar test for difference between predicted and actual population, p=0·18 for neural network versus p=0·0192 and p=0·0001 for clinician 1 and 2, respectively; table 4). In particular, there was a much higher probability that the neural network would correctly predict death. Both the surgeons and the neural network scored well when predicting survival, but this result is to be expected since 93% of the patients survived. The likelihood ratio, an index of the clinical usefulness of the prediction, was far greater for the neural network than for the clinicians.

Discussion

When dealing with cancer patients, most clinicians avoid trying to answer the question "How long have I got?". We are unable to predict with any degree of accuracy whether and when the individual patient will die, largely because of the failure of traditional statistical analysis of population databases to provide accurate outcome prediction for the individual patient.

Prediction of outcome for patients with colorectal tumours that are treated at an early or late stage is simple both clinically and mathematically, since by far the majority of patients with Dukes' A tumours survive, whereas most of those with established liver metastases rapidly succumb. Both clinicians and traditional statistics would correctly identify stage as being the most important variable for these patients. However, for Dukes' B and C tumours the situation is more complex, and traditional methods break down because of the influence of a potential multitude of clinical variables. Advances in computer processing speed and neural network theory have facilitated the application of neural networks to the non-linear analysis of complex data in many different settings, including diagnosis and prediction of outcome in medicine.[9–16]

Various other forms of artificial intelligence have also been used in medicine and surgery with varying degrees of success. Expert systems are programs that attempt to encode explicit representations of human expertise.[17] They are time-consuming to construct since they require

Survival period (months)	Overall accuracy (%) (95% CI)	Sensitivity (%)	Specificity (%)	Mortality (%)	Positive predictive value (%)	Negative predictive value (%)	Likelihood ratio (95% CI)
9	82 (72–90)	42	93	21	61	86	6·0 (1·5–21)
12	81 (70–91)	52	90	24	63	86	5·4 (1·7–16)
15	81 (70–91)	60	89	28	68	85	5·6 (2·0–16)
18	81 (70–91)	65	87	29	68	86	5·2 (2·0–13)
21	81 (70–92)	71	86	32	70	86	4·9 (2·0–12)
24	80 (69–91)	71	85	35	71	85	4·7 (2·0–11)

Table 3: **Performance of artificial neural networks trained and validated on data from institution I**

Predictor	Overall accuracy (%) (95% CI)	Sensitivity (%)	Specificity (%)	Mortality (%)	Positive predictive value (%)	Negative predictive value (%)	Likelihood ratio (95% CI)
Clinician 1	79 (71–87)	33	82	7	11	95	1·9 (0·56–6·5)
Clinician 2	75 (66–84)	67	76	7	16	97	2·7 (1·4–5·4)
Artificial neural network	90 (84–96)	67	92	7	36	98	8·2 (3·3–20)

Table 4: **Performances of clinicians and previously trained 12-month artificial neural network on data set from institution II**

protracted discussions with clinical experts and are fragile in operation. In addition, previous studies have shown that clinicians are not always aware of the important relations between variables that a neural network is able to detect.[18] In our study, the networks were more accurate than the experts, which implies that the complexity of the system may be beyond the analytical capabilities of a physician. Early work in Bayesian analysis has been used for prediction and diagnosis in medicine.[9] At that time the general belief was that a Bayesian classifier could produce good results only in conditions rarely found in practice.[20–22] Subsequent work has seen the application of Bayesian inference techniques to neural networks.[23]

In general, a neural network analysis is potentially more successful than traditional statistical techniques when the importance of a given prognostic variable is expressed as a complex unknown function of the value of the variable, when the prognostic impact of a variable is influenced by other prognostic variables, or when the prognostic impact of a variable varies over time.[24] These conditions are found in complex biological systems. Such systems have been extensively investigated in mathematics, physics, and theoretical biology.[25] For the purposes of analysis, each variable can be regarded as a single dimension in a multidimensional space. Traditional statistical techniques are particularly suited to the analysis of data with a low dimensional complexity and linear separation. In two dimensions, such data can be represented by the drawing of a straight line between two populations plotted on a Cartesian graph.[26] However, these traditional techniques are unable to provide such a division when the relation between variables is governed by a complex multidimensional non-linear function.[4] In these circumstances a neural network will provide a more accurate analysis.

There are several potential sources of error in our study. All the data were analysed retrospectively to permit the training and validation of the networks and to establish their ability to predict outcome for patients from a separate institution. However, the outcome of the 50 randomly selected patients used to validate each neural network was concealed from the computer. The outcome for patients from the second institution was not made known to the team involved in using the neural network until after the neural network had made its predictions. In fact, since the operation of the neural network is entirely automatic once input data are presented, there is no way to influence its prediction.

The proportion of deaths in the randomly selected patients from the second institution was only 7%, and the data on sensitivity must therefore be viewed with caution until a larger prospective series has been analysed. However, there were a substantial number of deaths in the larger training and validation set, and no significant adverse effect on the overall accuracy was apparent.

We accepted the accuracy of the data received from the separate institutions, which have their own internal audit systems. The parallel nature of the analyses performed by the neural networks enables them to accept a certain amount of inaccurate data without a serious effect on the predictive accuracy—a process known as graceful degradation. In contrast, the performance of expert systems is often unpredictable if input data are incorrect.

We discarded a variable from the study only if it was obviously unrelated to survival. This policy was justified on the grounds that we do not know which variables are the key factors, and there is some evidence[27] that the neural networks of the type used in this study are able to reduce automatically the contribution made by variables of low predictive value. Clinicians, however, are unable to detect accurately all the important relations within a complex system.[18] Interestingly, when the data from institution II were subjected to an effectively linear analysis by utilising a single-layer network, an overall accuracy of only 75% was achieved, which was similar to that achieved by the clinicians.

The clinicians were supplied with the data in tabulated form only and were not able to interview and examine the patients themselves before making their predictions. Diagnosis and prediction rely on the processing of information from all facets of the clinical situation, and the way the clinicians were required to perform in this study denied them the full use of their diagnostic skills.

In this study, artificial neural networks were able to provide much greater predictive accuracy for cause-specific death from colorectal cancer for individual patients than that provided by clinical judgment. This finding has implications for the selection of individual patients for adjuvant treatment irrespective of the population risk attributable to the tumour stage. The fact that a neural network trained in one institution was able to predict outcome accurately for another institution without retraining may facilitate comparative clinical audit. In addition, the ability accurately to predict outcome for patients treated by conventional methods may allow the effect of new treatments to be assessed within clinical trials without reliance on a large control population. However, a prospective study involving much larger data sets is required to confirm the findings of this preliminary study.

We do not wish to claim that neural networks are the answer for all complex data analysis. In fact, careful clinical judgment, especially during data preparation, is required to ensure that the neural network is producing meaningful results. There is also the ever-present danger, as with all mathematical methods of data generalisation, that the general relations discovered are not cause-effect relations. These relations can, however, be used to guide biomedical research directed at causal relations. As the complex nature of clinical prediction becomes apparent and the reductionist approach to prediction continues to disappoint, clinicians should be encouraged to review both their clinical paradigm and their method of analysis. Analysis of one or two variables will not permit accurate prediction in complex biological systems. Our recognition of the complexity of prediction does not mean that we have abandoned the desire for concise and simple relations between data and outcome, and we are mindful

that black-box approaches are ultimately unsatisfactory. So far, we have shown only that a neural network can discover a useful predictive relation. An important part of our future work is to find out whether we can re-express the predictive relation in a form that is more intelligible to the clinician, involving fewer key factors rather than the large number of variables used in this study. To this end, symbolic rule extraction is a promising technique that will be investigated.[28] The multilayer perceptron model with back propagation learning as used in this study may not be suitable for rule extraction. In this event, a different neural network model better suited to rule extraction, such as adaptive resonance theory mapping neural network (ARTMAP),[29,30] can be used.

In the future, by accepting the shift from simple and linear to complex and non-linear and by applying analytical tools such as artificial neural networks, clinicians may well find the answers that they seek.

This study was supported by a Northern and Yorkshire regional research fellowship (PJD) and by a British Oncological Association health research grant.

References

1 Fisher ER, Sass R, Palekar A, et al. Dukes' classification revisited. *Cancer* 1989; **64:** 2354–60.

2 Jass JR, Love SB, Northover JMA. A new prognostic classification of rectal cancer. *Lancet* 1987; i: 1303–06.

3 Newland RC, Dent OF, Lyttle MNB, et al. Pathologic determinants of survival associated with colorectal cancer with lymph node metastasis. *Cancer* 1994; **73:** 2076–82.

4 Rumelhart D, McClelland J, eds. Parallel distributed processing: explorations in the microstructures of cognition. Cambridge, MA: MIT Press, 1986.

5 Bishop CM. Neural networks for pattern recognition. Oxford: Clarendon Press, 1995.

6 Stuttgart Neural Network Simulator V4.0. Institute for Parallel and Distributed High Performance Systems, University of Stuttgart, Germany, ftp.informatik.uni-stuttgart.de.

7 Rumelhart D, Hinton G, Williams R. Learning representations by back propagating errors. *Nature* 1986; **323:** 533–36.

8 Altman DG, Bland MJ. Diagnostic tests 2: predictive values. *BMJ* 1994; **309:** 102.

9 Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann Intern Med* 1991; **115:** 843–48.

10 Baxt WG. Application of neural networks to clinical medicine. *Lancet* 1995; **346:** 1135–38.

11 Doye HR, Dvorchik I, Mitchell S, et al. Predicting outcomes after liver transplantation. *Ann Surg* 1994; **219:** 408–15.

12 Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat* 1992; **22:** 285–93.

13 Ravdin PM, Clark GM, Hilsenbeck SG, et al. A demonstration that breast cancer recurrence can be predicted by neural network analysis. *Breast Cancer Res Treat* 1992; **21:** 47–53.

14 McGuire WL, Tandon AK, Allred DC, et al. Treatment decisions in auxillary node-negative breast cancer patients. *J Natl Cancer Inst Monogr* 1992; **11:** 173–80.

15 DeLaurentis M, Ravdin PM. Survival analysis of censored data: neural network analysis detection of complex interactions between variables. *Breast Cancer Res Treat* 1994; **32:** 113–18.

16 Clark GM, Hilsenbeck SG, Ravdin PM, et al. Prognostic factors: rationale and methods of analysis and integration. *Breast Cancer Res Treat* 1994; **32:** 105–12.

17 Clancey WJ, Shortliffe EH, eds. Readings in medical artificial intelligence: the first decade. Reading, MA: Addison Wesley, 1984.

18 Baxt WG. A neural network trained to identify the presence of myocardial infarction bases some decisions on clinical associations that differ from accepted clinical teaching. *Med Decision Making* 1994; **14:** 217–22.

19 Horrocks JC, McCann AP, Staniland JR, et al. Computer aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases. *BMJ* 1972; ii: 5–9.

20 Engle RL, Davis BJ. Medical diagnosis: present past and future I, present—concepts of the meaning and limitations of medical diagnosis. *Arch Intern Med* 1963; **112:** 108–15.

21 Engle RL. Medical diagnosis: present past and future III, diagnosis—the future including a critique on the use of electronic computers as diagnostic aids to the physician. *Arch Intern Med* 1963; **112:** 126–39.

22 Croft DJ, Machol RE. Mathematical models in medical diagnosis. *Ann Biomed Engin* 1974; **2:** 69–89.

23 MacKay DJC. Bayesian methods for backpropagation networks. In: Domay E, van Hemmen JL, Schulten K, eds. Models for neural networks III. New York: Springer Verlag, 1994.

24 De Laurentis M, Ravdin P. A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Lett* 1994; **77:** 127–38.

25 Kaplan D, Glass L, eds. Understanding nonlinear dynamics. New York: Springer Verlag, 1995.

26 Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995; **346:** 1075–79.

27 Hartman EJ, Keeler JD, Kowalski JM. Layered neural networks with gaussian hidden units as universal approximations. *Neural Computation* 1990; **2:** 210–15.

28 Ma Z, Harrison RF, Cross SS. Explanation by general rules extracted from trained multi-layer perceptrons. Automatic control and systems engineering research report no 650, University of Sheffield, Sheffield, 1996.

29 Carpenter GA, Tan A. Rule extraction, fuzzy ARTMAP and medical databases. *Proc World Congr Neural Networks* 1993; **I:** 501–06.

30 Downs J, Harrison RF, Kennedy RL, Cross SS. Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks. *Artif Intellig Med* 1996; **8:** 403–28.