

Title: The use of Artificial Neural Networks for risk prediction following Carotid endarterectomy

Authors: G Kuhan, DN Davis*, CN McCollum⁺, PT McCollum

Departments:

Academic Vascular Unit, Hull Royal Infirmary,

Anlaby Road, Hull, HU3 2JZ

*Department of Computer Science, University of Hull

⁺Academic Surgery Unit, South Manchester University Hospital

Category: Original article

Correspondence to:

Professor PT McCollum

Academic Vascular Unit, Hull Royal Infirmary,

Anlaby Road, Hull, HU3 2JZ, United Kingdom.

Tel No- 01482 674704

Fax- 01482 675665

Email- peter.mccollum@virgin.net

Short title: Artificial Neural Networks for risk prediction following Carotid endarterectomy

Introduction

Carotid endarterectomy (CEA) is emerging as the commonest arterial procedure performed by vascular surgeons. The efficacy of CEA in stroke prevention has been proven by randomised controlled trials¹⁻³. Comparative audit of outcome of vascular units and surgeons will ensure provision of the highest quality of care⁴. Risk models have been used to compare risk adjusted outcomes⁵⁻⁷. Risk models identify higher risk groups of patients- which can be used to counsel patients before the procedure-and aid in comparison of risk adjusted outcomes on a regular basis⁸. However very few models accurately predict individual patient's risk, especially when the outcome event rates are low. Estimating risk for a high risk group of patients often has a wide confidence interval, thus the mean of the distribution cannot be relied on for estimating an individual patient's risk. Linear models assume a linear relationship between the variables and the outcome; however data in most instances have a complex relationship between the variables and outcome. Furthermore, linear models often fail if missing or "noisy" data are encountered. Advances in computer processing speed and neural network theory has enabled the use of non-linear models for the analysis of complex data for diagnosis and outcome prediction in medicine⁹⁻¹⁷.

Neural networks are distributed, adaptive non linear learning machines based on a information processing model that emulates human brain activity. The basic unit, the processing element (PE), has input and output parameters, similar to a neuron. Each PE performs a relatively simple task but a combination of several PEs, connected by weighted links, work simultaneously, to perform parallel processing, and produce an incredible processing unit, similar to the human brain. One neural network learning model is a variation on trial and error termed back propagation. An error obtained by

comparing the output of the network with the desired is feed back to adjust the weights connecting the processing elements^{18,19}. The multi layer perceptron (MLP) is the most commonly used network for classification^{20,21}. In addition to the input and output layers, variable number of hidden layers can be added in MLPs. Generalized feed forward networks (GFFN) are similar to MLPs but the connections can jump over one or more layers thus enabling a faster more efficient training. Radial basis function networks (RBF) view complex data at a higher dimension, in an attempt to make classification easier²²⁻²⁴.

Aims

The aims of the study were to develop neural network models that accurately predict outcome for individual patients and to compare the predictive accuracy of various types of neural network models (MLP,GFFN, RBF).

Methods

Training set

Prospective and retrospective data on 839 CEAs carried out over a period of 5 years from 2 clinical units were used for training the ANNs. The data set was simplified to 15 possible risk factors (Table 1). Details of the data collection methods are described in a previous study⁸. The sample size was considered adequate for an error rate of 0.02 for the training set²⁵. The outcome end points were the occurrence of stroke or death within 30 days of the operation. Twenty four percent (200 CEAs) of the training data were randomly selected for cross validating the trained model. The training set was randomised 10 times, with different cross-validation cuts, to produce 10 training and cross-validation sets.

Validation set

Prospective data on 702 CEAs from a further 2 vascular units collected over the same period of time were available to test the developed models. The data set was simplified to the same 15 variables given in table 1.

Coding fields

Data values for the variables were converted to numerical values which lie between 0 and 1.

Design of ANN

MLPs were used initially to develop the predictive model since the analytical power of these networks are well established²⁶. Neural Solution 4.2 (Neural Dimension, Florida, USA) was used to develop MLP, GFFN and RBF network models. The following parameters of the network were systematically changed - number of hidden layers (0-6), number of processing elements in each of the hidden layers, type of axon used, learning rule, step size and momentum- to try and improve the performance. The

following specific parameters in the RBFs were altered to increase the performance of the network- cluster centers, competitive rule and distance metric. All networks were trained to a maximum of 1000 epochs.

Training the ANNs

The training rule used back propagated error to reduce the overall mean squared error (MSE). An oscillating or rising learning curve over the training cycle indicates poor learning.

Cross validation

The trained networks performance was assessed on random sample of 15% of the training data which was not used for training. The learning curve of both cross validation and training sets should decrease with number of epochs to produce better performance. The minimum mean squared error (MSE) was compared for each network for training and cross validation sets.

Statistical comparison

The predictive accuracy of the networks to classify patients with strokes or deaths after CEA was assessed by simple classification matrixes. The sensitivity, specificity, positive predictive value, negative predictive value and the likelihood ratios were compared for each network.

Results

The observed stroke or death rates was 3.9 % (/839) and 4.4% (/702) for the training and the validation set respectively.

The lowest minimum MSE was observed with the 0 hidden layer MLP, 1 hidden layer GFFN and 1 hidden layer RBF network (Figure 1). Increasing the number of hidden layers did not reduce the minimum MSE in the training and cross validation sets. The 95% confidence intervals of the minimum MSEs became wider with increase in the number of hidden layers.

The best accuracy of prediction in the training set was obtained with MLP 1, 2, 3 hidden layers and GFFN 1, 2 hidden layers (Figure 2). However, the predictive accuracy did not increase with increasing the number of hidden layers. MLPs, GFFN and the RBFs produced good training models and out performed the linear regressor (0 hidden layer MLP) for predictive accuracy in the training set.

The accuracy of the model to classify strokes and deaths were poor in all networks when used on the validation set (Figure 3). The best predictive accuracy was obtained with RBF 0 hidden layer. This significantly out performed all other networks (Mean predictive accuracy of 16.80 for strokes or death and 84.34 for non stroke or death). Sensitivity, positive predictive value and likelihood ratios were poor for all the networks. However RBFs performed better than the MLPs, GFFNs and the linear regressor (Figures 4-6).

Discussion

For CEA to be worthwhile the long term benefit (stroke prevention) should outweigh the immediate operative risk (stroke or death). The benefit of CEA for patients with symptomatic high grade carotid lesions (70-99%) has been proven using large randomised controlled trials^{1,2}. The benefit for patients with asymptomatic high grade lesion is currently being evaluated²⁷. Although CEA reduces the risk of stroke by 50% compared to best medical treatment for patients with high grade lesions, over three years only 20% of these patients will have a stroke if not operated on. Therefore, surgery is not beneficial in nearly 80% of patients despite them having a symptomatic high grade carotid lesion. CEA can be harmful in some of these patients, especially if the operative risk is unexpectedly high²⁸. Therefore individual patient risk rather than surgeons own observed operative risk is important for CEA. Existing risk models successfully identify a high risk group of patients, but the mean predicted risk is often associated with wide confidence intervals^{8,29}. In the current study using various neural network models the accuracy of individual patient risk was evaluated on data which was blinded and not used for training. The data sets from 2 further institutions were sought only after developing various risk models. Despite using advanced computer technology the percentage predictive accuracy, sensitivity, positive predictive value and the likelihood ratio were poor in the validation set. The results of this study suggest the developed neural models do not generalize to previously unseen data. Despite this negative result, there were several salient points to learn.

The training set consisted of 839 CEAs. The risk variables were reduced to 15 risk factors from the initial 67. Univariate analysis, previous studies and clinical relevance were used to guide the risk factor selection⁸. The percentage prevalence of missing

data for the selected risk factors was 1.4%. Missing values are inevitable in administrative data and Neural Networks handle missing data, to an extent better than traditional statistical methods. The computational nature of the neural net model compensates for missing and “noisy” data by a process known as graceful degradation. The most commonly encountered problem in any automated model development is “over fitting” the model to the trained data set. A random sample of 24% (200 CEAs) of the training data was used to cross validate the training model. Cross validation is a well recognised statistical method of reducing the statistical bias of learnt models towards the training data³⁰. It prevents the learnt model becoming too specific (idealised for the training data) and enables better predictive accuracy when the model is exposed to new data.. Weights of the neural model were selected through halting the network training when the mean squared error (MSE) was low for both training and cross validation data sets. Furthermore, 10 randomised training sets were produced to reduce the probability of developing a biased neural model by chance.

It was possible to develop good predictive models on the training set using MLP and GFFN networks. The desired MSE of 0.02 was not achieved. Although nets with hidden layers were more accurate than nets with no hidden layers, increasing the number of hidden layers did not reduce the MSE or increase the accuracy of prediction. Not only did increasing the hidden layers slow the time to train, it increased the error of prediction. The neural network models trained better than the linear predictor. Multiple variables have a complex relationship with the outcome, hence the simpler linear models fail when they encounter such data (GANESH _ NOTE that LINEAR models can be used in any order of dimensions of space! – remove this after reading!) Tremendous development has been made in Neural Network theory and processing

speed enabling them to adapt to the data at various levels of complexity. Connections in GFFNs jump from one layer to another unlike the inter-layer connectivity of MLPs, thus decreasing the learning time. RBF networks add a Gaussian distribution to the input data thus increasing the dimensionality of the classifier over the data. RBFs also have the ability to cluster data without supervision. They can determine for themselves which data is alike since each PE represents a local cluster of patients who are classified similarly over some level of data dimensionality. RBFs in our study trained poorly compared to the other networks but had a better predictive accuracy in the validation set. The RBFs ability to see data at this increased level of dimensionality and the ability to self-cluster data aided to train a model with a high predictive accuracy, albeit with less over-fitting.

Overall accuracy of all models was poor in the validation set. The data used for training the models were routinely collected administrative data. The data volume might not be sufficient to identify the complexity of predicting individual patient risk. Consider a MLP with 15 input nodes, 5 hidden layer nodes and 1 output node, resulting in 80 weighted connections. One heuristic for training this type of net suggests that there should be at least 10 training examples of each observed class per connection weight. This means that we need at least 800 examples of each class to accurately train the network for this application. This volume of data is not yet available. Thirdly, the data used for training and validation might have some inherent difference that is not explained in terms of the statistical analyses associated with this study. Finally failure to find a useable solution in this study does not mean the application is not amenable to neural solutions. Further investigations into the data and the use of neural, fuzzy neural and rough-set based models are on-going.

Acknowledgement

The study was funded by “freemasons” research fellowship, Royal College of Surgeons of England.

Figure 1 Minimum Mean Squared Errors (MSE) for training and cross validation sets for the Networks

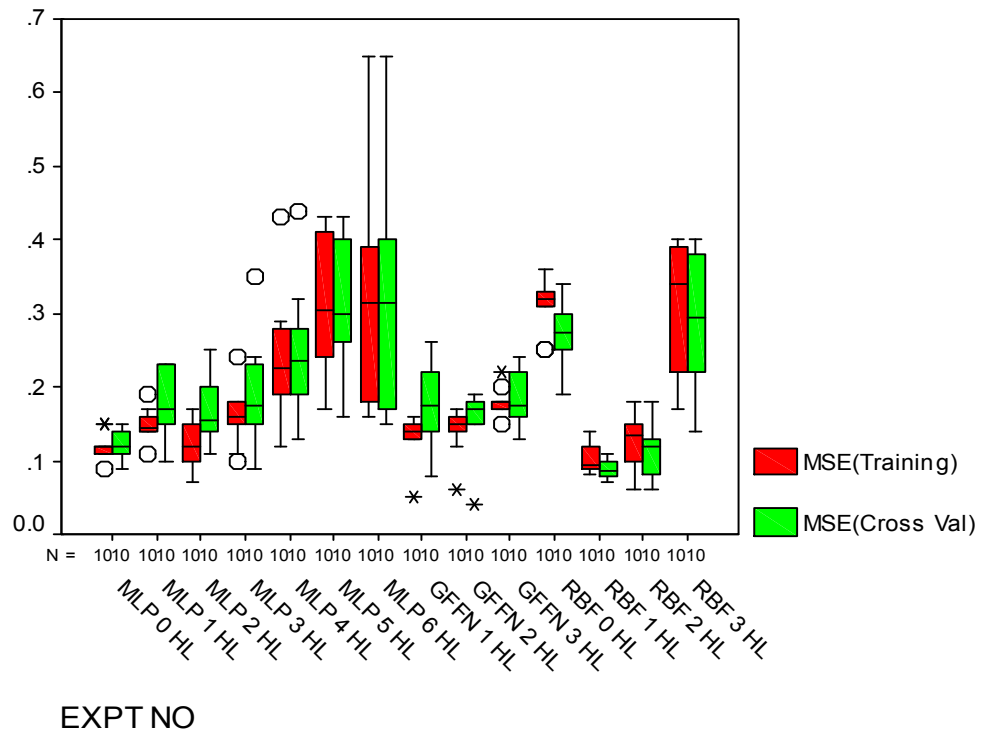


Figure 2 Mean percentage of accuracy of prediction for strokes or deaths and non strokes and death in the training set

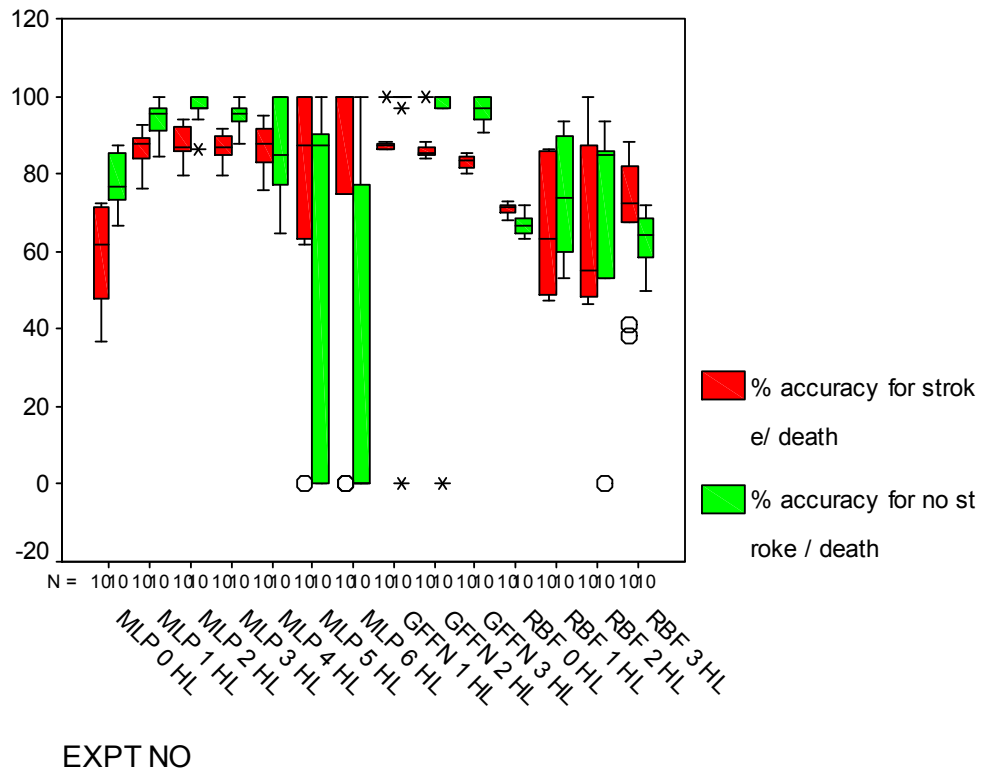


Figure 3 Mean percentage of accuracy for strokes or death and non stroke or deaths in the validation set

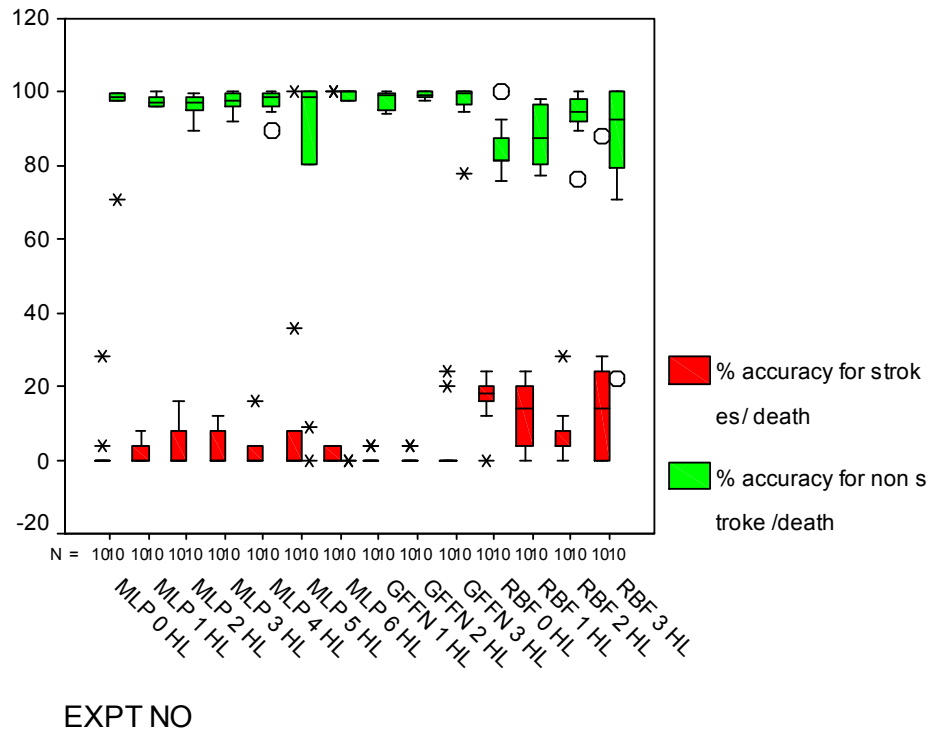


Figure 4 Mean sensitivity and specificity on the validation set

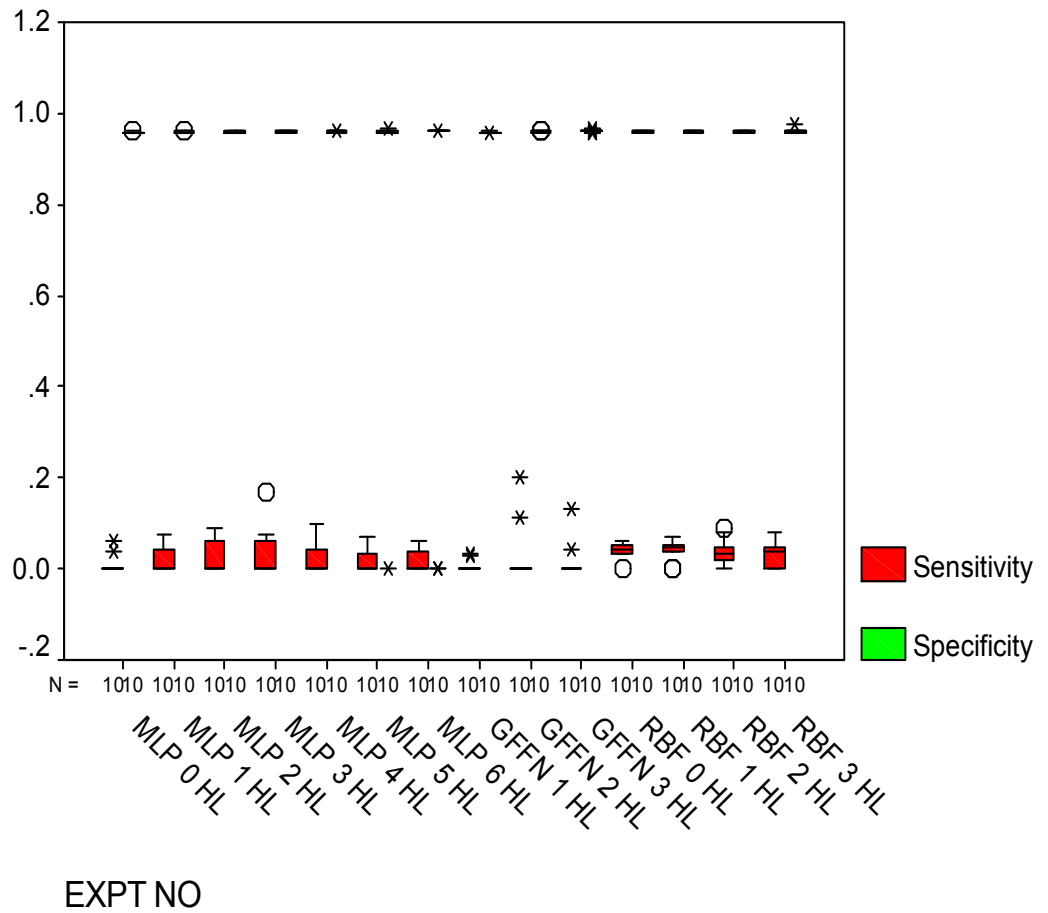


Figure 5 Mean PPV and NPV on the validation set

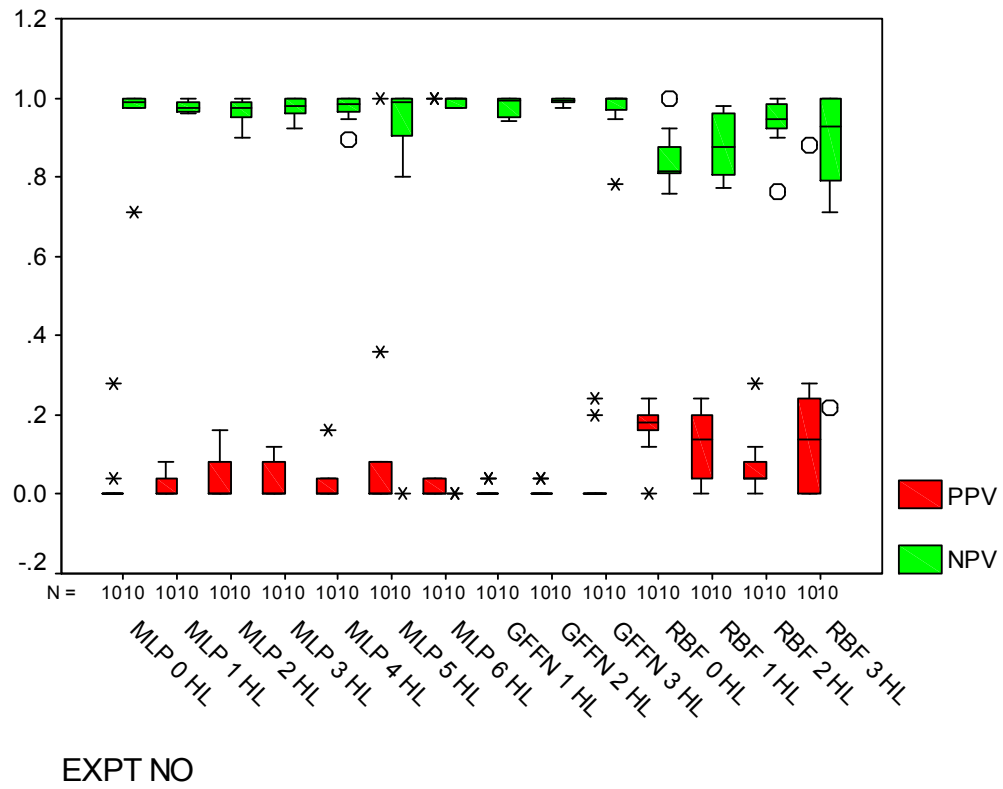


Figure 6 Mean prevalence and likelihood ratios on the validation set

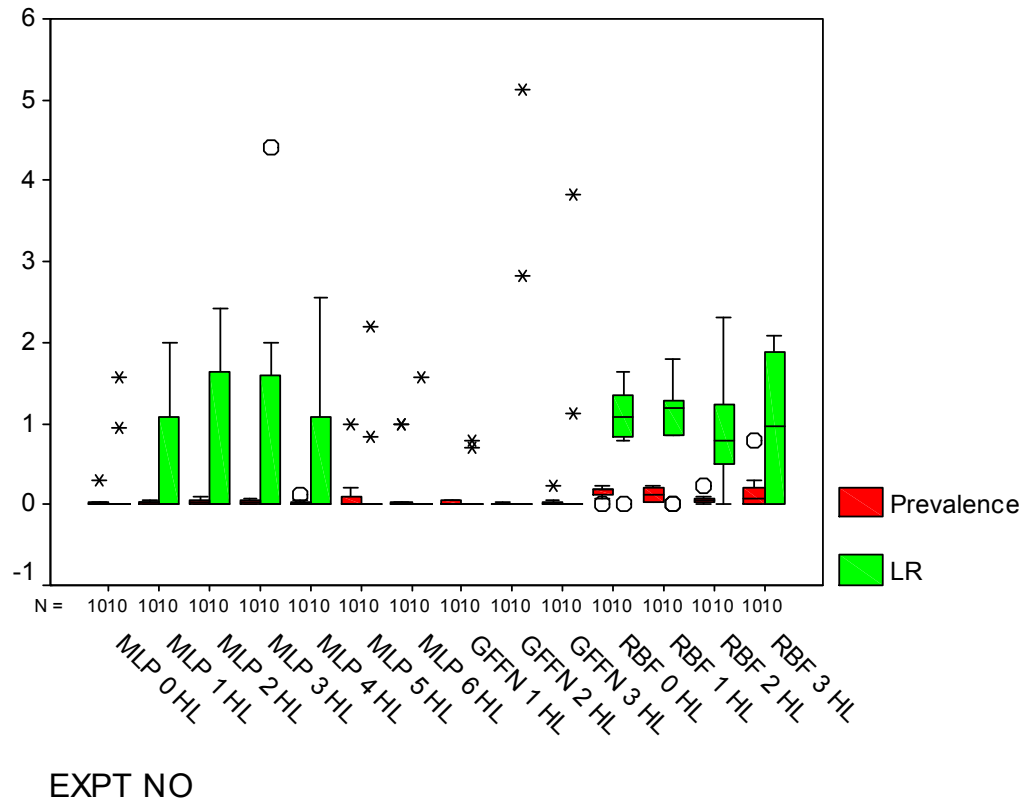


Table 1: Selected risk factors for the model

Age	Respiratory disease
Sex	Side of operation
Hypertension	Shunt
Heart disease	Patch
Diabetes	ASA grade
Stroke	Surgeon
Renal failure	Vascular unit
Contra lateral internal carotid artery occlusion	

Reference List

1. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998;351(9113):1379-87.
2. Barnett HJ, Taylor DW, Eliasziw M, Fox AJ, Ferguson GG, Haynes RB et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N.Engl.J.Med* 1998;339(20):1415-25.
3. Mayberg MR, Wilson SE, Yatsu F, Weiss DG, Messina L, Hershey LA et al. Carotid endarterectomy and prevention of cerebral ischemia in symptomatic carotid stenosis. Veterans Affairs Cooperative Studies Program 309 Trialist Group. *JAMA* 1991;266(23):3289-94.
4. NHS management executive and Department of health. Clinical governance: Quality in the newNHS. 1999.
5. Bernstein AD, Parsonnet V. Bedside estimation of risk as an aid for decision-making in cardiac surgery. *Ann.Thorac.Surg.* 2000;69(3):823-8.
6. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma* 1989;29(5):623-9.
7. Copeland GP. The POSSUM system of surgical audit. *Arch.Surg.* 2002;137(1):15-9.
8. Kuhan G, Gardiner ED, Abidia AF, Chetter IC, Renwick PM, Johnson BF et al. Risk modelling study for carotid endarterectomy. *Br.J Surg.* 2001;88(12):1590-4.

9. Bottaci L, Drew PJ, Hartley JE, Hadfield MB, Farouk R, Lee PW et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet* 1997;350(9076):469-72.
10. Buzatu DA, Taylor KK, Peret DC, Darsey JA, Lang NP. The determination of cardiac surgical risk using artificial neural networks. *J Surg.Res.* 2001;95(1):61-6.
11. Droste K, Bollschweiler E, Waschulzik T, Schutz T, Engelbrecht R, Maruyama K et al. Prediction of lymph node metastasis in gastric cancer patients with neural networks. *Cancer Lett.* 1996;109(1-2):141-8.
12. Izenberg SD, Williams MD, Luterman A. Prediction of trauma mortality using a neural network. *Am.Surg.* 1997;63(3):275-81.
13. Naguib RN, Adams AE, Home CH, Angus B, Smith AF, Sherbet GV et al. Prediction of nodal metastasis and prognosis in breast cancer: a neural model. *Anticancer Res.* 1997;17(4A):2735-41.
14. 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(3):285-93.
15. Wong LS, Young JD. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks. *Anaesthesia* 1999;54(11):1048-54.
16. Doyle HR, Dvorchik I, Mitchell S, Marino IR, Ebert FH, McMichael J et al. Predicting outcomes after liver transplantation. A connectionist approach. *Ann.Surg.* 1994;219(4):408-15.
17. Turton EP, Scott DJ, Delbridge M, Snowden S, Kester RC. Ruptured abdominal aortic aneurysm: a novel method of outcome prediction using neural network technology. *Eur.J Vasc.Endovasc.Surg.* 2000;19(2):184-9.

18. Rumelhart D, Hinton G, Williams R. Learning representations by backpropagation. *Nature* 1986;323:533-6.
19. Werbos PJ. Backpropagation through time: what it does and how to do it. *IEEE Eng Med Biol.Mag.* 1990;78(10):1550-60.
20. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346(8982):1075-9.
21. Hornik K, Stinchcombe M, White H. MLPs are universal approximators. *Neural Networks* 1989;2:359-66.
22. Broomhead DS and Lowe D. Radial basis functions, multivariable functional interpolation and adaptive networks. 1988. Malvern, Royal signal and radar establishment.
23. Haykin S. *Neural networks: a comprehensive foundation*. New York: Macmillan; 1994.
24. Park J, Sandberg I. Universal approximation using radial basis function networks. *Neural Comput.* 1989;3:303-14.
25. Baum EB, Haussler D. What size net gives valid generalization. *Neural Comput.* 1989;1:151-60.
26. Bishop C. *Neural networks for pattern recognition*. Oxford: Clarendon Press; 1995.
27. Halliday AW, Thomas D, Mansfield A. The Asymptomatic Carotid Surgery Trial (ACST). Rationale and design. Steering Committee. *Eur.J.Vasc.Surg.* 1994;8(6):703-10.
28. Rothwell PM, Slattery J, Warlow CP. A systematic review of the risks of stroke and death due to endarterectomy for symptomatic carotid stenosis. *Stroke* 1996;27(2):260-5.

29. Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. European Carotid Surgery Trialists' Collaborative Group. *Lancet* 1999;353(9170):2105-10.
30. Schumacher M, Hollander N, Sauerbrei W. Resampling and cross-validation techniques: a tool to reduce bias caused by model building? *Stat.Med* 1997;16(24):2813-27.