Breast Cancer Survival Prediction using Artificial Neural Network

Venkatesan P† and Suresh M.L. ††

† Department of Statistics, Tuberculosis Research Centre, ICMR, India, Chennai
†† Research Scholar, Sathyabama University, India, Chennai

Abstract
Breast cancer survival prediction is an important step in the complex decision process. The present study investigates the effects of prognostic variables on the risk of breast cancer failure/survival using feed forward neural network. The neural network was trained and tested using six hundred and fifty breast cancer patients. 15 prognostic variables were used in this study as features for the input vectors. Multivariate analysis on survival was performed using a recursive partitioning namely classification and regression trees (CART). The results show that the accuracy of artificial neural network (ANN) for the survival prediction was better than regression based approaches.

Key words:
ANN, Breast cancer, MLP, CART, logistic regression

1. Introduction

Artificial neural networks are a class of pattern recognition methods that have been successfully implemented in data mining and prediction problem in a variety of fields [1-4]. ANN is preferred to solve these problems, because of their parallel processing capabilities, as well as decision-making abilities. In recent years ANN has been widely used in biological and medical research such as cancer outcome, survival prediction and breast cancer diagnosis etc., [5-14]. Neural network generalizes from the input data to patterns inherent in the data, and uses these patterns to make predictions. Prediction is a method to forecast the outcome of a specific situation, using input information obtained from the prognostic variables, which describes the situation. When one is interested in future prognosis of therapy response, the problem is considered as survival prediction. The challenge of training the neural network using the clinical data is to extract information hidden in the risk factors at the time of forecast. Not all biological factors influence the course of a disease as measurable factors. In addition, the effects of hidden factors increase the noise in the data resulting in poor prediction. Since many factors are correlated with each other, it results in over fitting [15].

Breast cancer is one of the most common cancers in women [16]. Survival estimations are generally performed using non-parametric techniques. For analysis of prognostic factors of patients with cancer, models such as CART and the logistic regression model are widely used [17-19].

The regression tree is constructed by repeatedly splitting the data using a simple rule on dependent variable. At each split the data is divided into two different groups. The splitting procedure is applied to each group again and again until a large tree is grown. Then the tree is pruned back to the required size. The objective is to divide the variables into homogeneous groups so that the tree is as small as possible.

ANN has shown to be a powerful tool for analyzing the data when there are non-linear interactions between the input and the information to be predicted. Recently, neural networks are compared with other statistical methods for survival prediction [20-22]. The results are applied to breast cancer patients: follow up intervals, in order to obtain survival prediction using clinical – pathological data. Different network topology and learning parameters are investigated to obtain the best prediction accuracy.

2. Breast Cancer Data

The data consists of patients registered for suspected breast cancer tumor from 2000 to 2003 in the Cancer institute, Chennai, India. Six hundred and fifty patients who followed for 5 years treatment or until death occurred from breast cancer were considered for this study. The covariates comprises of age of the patient (AGE), age at menarche (AMR), age at marriage (AMA), number of children (NOC), breast feed (BFD), age at first child birth (FCB), family history (FAH), age at menopause (AMO), tumor size (cm)(TSZ), nuclear grade (low, medium, high)(NGR), tumor stage (2-4)(TSG), chemotherapy (CT), radiotherapy (RT), hormone therapy (Ht) and surgery (SGY). Staging is the overall staging index with 3 grades according to the severity of the tumor, assessed by the doctors at the time of admission. Every patient on this database was followed up for at least 5 years from the date of admission. Some of the prognostic variables are binary. The baseline demographic and disease characteristics of the patients for the three stages are given in Table 1 and 2.

Manuscript received May 5, 2009
Manuscript revised May 20, 2009

Table 1: Mean (SD) of covariates according to stage

Variable	Stage II n(170)	Stage III n(348)	Stage IV n(132)	All n(650)
AGE	49.3(12.8)	49.0(11.2)	47.2(11.4)	48.7(11.7)
AMA	19.5(5.3)	19.4(4.8)	19.6(5.3)	19.5(5.0)
FCB	21.2(6.4)	20.2(7.6)	20.4(7.8)	20.5(7.4)
NOC	2.4(1.16)	2.43(1.53)	2.33(1.49)	2.5(1.9)
TSZ	3.90(1.15)	7.9(3.0)	8.1(2.9)	6.9(3.2)

Table 2: Distribution of covariates according to stage

Variable		Stage II n(170)	Stage III n(348)	Stage IV n(132)	All n(650)
		%	%	%	%
AMR (<13)		16.5	10.3	18.9	13.7
AMO (>54)		5.9	3.7	1.5	3.8
NGR	I	26.5	18.4	19.7	20.8
	II	37.6	33.6	47.7	37.5
	III	35.9	48	32.6	41.7
FAH (yes)		8.8	6.0	6.8	6.9
BFD (yes)		92.9	85.1	87.9	87.7
CT (yes)		91.2	94.8	89.4	92.8
RT (yes)		92.4	94.8	72	89.5
HT (yes)		79.4	68.4	65.9	70.7
SGY (yes)		83.5	71.6	14.4	63.1

3. Regression Models

3.1 Logistic Regression

Binary logistic regression is one of the most commonly used statistical techniques when the dependent is dichotomous variable and the independent variables are continuous, categorical or both. In linear regression, the relationship between the independent variables and dependent variables is linear. But in logistic regression (LR) the relation between the variables is not assumed linear. It estimates the probability that the outcome variable assumes certain value, rather than estimating itself. The model is given by

$$\log\left(\frac{p}{1-p}\right) \Rightarrow x_0 + \sum_{i=1}^{n} a_i x_i \qquad (1)$$

where p = p (Y=1 / X).

3.2 CART

Classification and Regression Trees (CART) are widely used in machine learning applications and it is best suited for the analysis of complex data's. CART analysis is a tree building technique, which is unlike traditional data analysis methods. It is suited for generation of clinical decision rules. CART is able to uncover complex interactions between predictors, which may be difficult to uncover using traditional multivariate techniques. CART is a non-parametric technique, which selects the variables and their interactions, which are important in determining the dependent variable. If the dependent variable is continuous CART produces regression tress, if the variable is categorical CART produces classification trees. CART uses an accurate set of data classifiers by uncovering the predictive structure of the survival problem.Fig1. gives the CART diagram for the breast cancer data. Fig2 represents the relative importance of covariates.

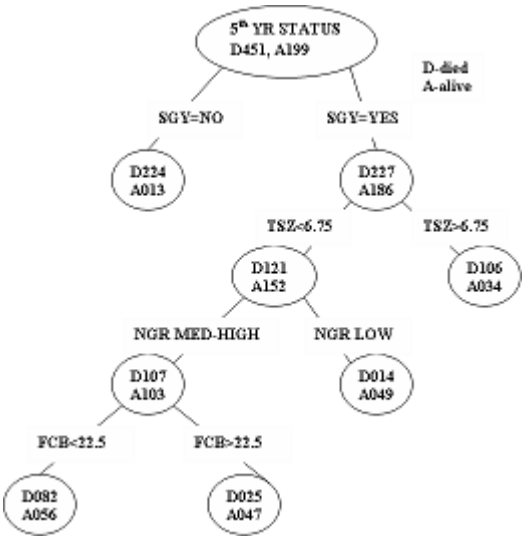

Fig.1 CART generated 5th year survival prediction.

CART is a rule based method that generates a binary tree structure through a binary recursive partitioning in which each parent node is split into two children nodes accor ding to yes/no answer of the predictors [23]. Purity function uses a single variable to split the tree. A split is selected such that the data in the child node is purer than the data in the parent node. Each node is split further based on the independent rules to find the threshold among the descriptive variables at the node of all dimension and they predict the training samples with least error.

CART has number of advantages over other classification methods including multiple logistic regressions. CART is non parametric and can handle numerical data that are

highly skewed or multi modal. CART can also handle categorical prediction with both ordinal or non ordinal structure and missing variables. When primary splitting variable is missing for an individual, that observation is not discarded, instead a surrogate splitting variable is selected within the data set similar to the primary splitter. The program uses the best available information for the missing value. This advantage is not there in tradition multivariate regression modeling where the missing values are generally discarded.

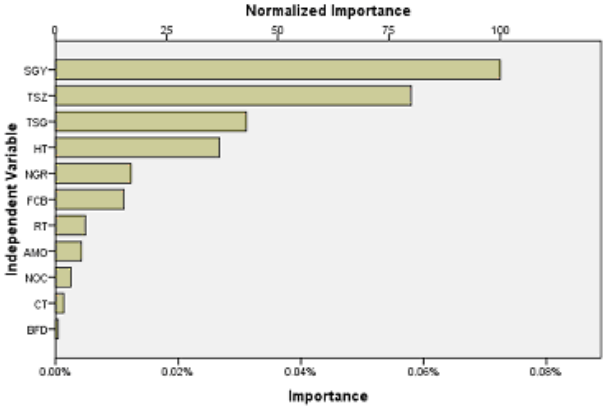

Fig-2 Normalized Relative Importance of covariates

CART analysis consists of four basic steps. First step is the Tree Building in which tree is built using recursive splitting of nodes. Second step is stopping Tree Building process. The tree building process goes on until it is impossible to continue. Thus maximal tree is generated which over fits the information contained in the learning data set. Third step is tree pruning. A technique known as cost-complexity pruning is used. It depends on complexity parameter α, which is gradually increased during the training process. Beginning at the terminal node, the child nodes are pruned such that predicted misclassification is less than α times the change in tree complexity. As α is increased more and more nodes are pruned resulting in simple trees. Fourth step is optimal tree selection. The maximal tree always fits the learning data set with higher accuracy than other trees. The tree, which fit the information in the learning data set but does not over fit the information, is selected among the sequence of pruned trees.

4. ANN

Multilayer perceptron (MLP) network models are one of the most popular ANN architectures used in most of the research applications in medicine, engineering and other applications [24]. In MLP the weighted sum of the inputs and bias term are passed to activation level through a transfer function to produce the output, and the units are arranged in a layered feed-forward topology called Feed ward neural Network (FFNN). Supervised learning is achieved by an iterative adjustment of the network connection weights in order to minimize an error function, computed over the training cases. The schematic representation of FFNN with 15 covariates and 9 hidden nodes and one output layer is given in fig (3). An ANN has 3 layers: input layer, hidden layer and output layer. The hidden layer vastly increases the learning power of the MLP. The output layer had one output node representing the survival of each patient at the end of the 4^{th} or 5^{th} year. The transfer or activation function of the network modifies the input to give the desired output. The choice of the number of hidden layers, hidden nodes and type of activation function plays an important role in model construction [25-28]. Back propagation error correction method is used for breast cancer survival prediction.

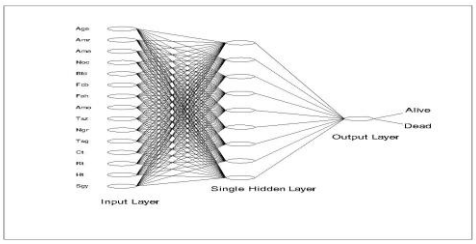

Fig-3 Survival prediction using FFNN

5. Methods

A multiplayer perceptron model of ANN using sigmoid function was constructed. The multi layer feed forward neural network was trained with back propagation algorithm for minimizing the error in each training epoch. To reduce the risk of over fitting and to test the networks generalizing ability, the dataset was divided into training set and testing set. The commonly used 2/3 and 1/3 partitions were adopted for the training and testing [29]. The training data was used to train the model and the test data was used to test the measure the performance of the network. The prognostic variables were fed in the network and the output was calculated using

$$y_k = f(\sum_{h=1}^{H} w_{hk} f(\sum_{i=1}^{I} w_{ih} x_i)) \quad (4)$$

Where f is the sigmoid function given by

$$f(x) = \frac{1}{1+e^{-x}} \quad (5)$$

Whose output lies between 0 and 1.In (4), x_i's are input values, w_{ih} are the weights from input to the hidden layer. After calculating the output value, the output is compared with the target value and the difference is used to determine the error. The error is calculated using

$$E(w)=\frac{1}{2}\ t_k - y_k\ ^2 \quad (6)$$

Where t_k and y_k are target and the output values. To minimize the error the weight adjustment is calculated using $\Delta w = -\eta \frac{\partial E}{\partial w}$ (7)

Where η is the learning rate.

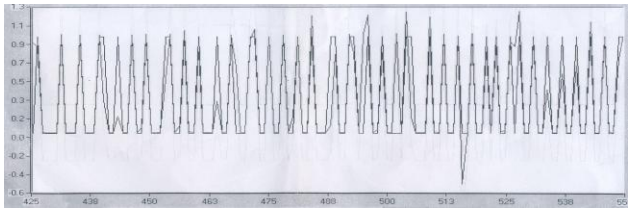

Fig-4 Time series graph for the actual and predicted survival

The Neural network software NeuNet Pro was used to fit the ANN model. The sensitivity is the probability of correctly predicting death, specificity is the probability of correctly predicting the survival and the accuracy is the percentage of correct predictions. The survival prediction error is given in table 3 and 4. The comparative survival prediction for 4th year and 5th year status for the regression models and the neural network is given table 5.

Table 3: Performance measures (4th year status)

↓ Actual/Predicted →	Dead	Alive	Total	Prediction Error (%)
Dead	**103**	**20**	**123**	**16.26**
Alive	**08**	**86**	**94**	**08.51**
Total	**111**	**106**	**217**	**12.90**
Actual Error (%)	**7.21**	**18.87**	**12.90**	

Table 4: Performance measures (5th year status)

↓ Actual/Predicted →	Dead	Alive	Total	Prediction Error (%)
Dead	**114**	**22**	**136**	**16.18**
Alive	**03**	**78**	**081**	**03.70**
Total	**117**	**100**	**217**	**11.52**
Actual Error (%)	**2.56**	**22.00**	**11.52**	

Table 5: Comparitive survival prediction of Regression models and NNW

Model/status	LR	CART	NNW
4th year	**78.0**	**74.5**	**87.16**
5th year	**80.0**	**78.2**	**88.48**

6. Discussion

The advantage of the neural networks lies in their ability to process nonlinear relationships. Because of the clinical complexity and pathologic heterogeneity of cancer, correct identification of patients with active disease is unlikely to depend on the presence of absence of a single defining feature. Hence, it is not surprising that standard linear statistical methodologies are relatively inadequate solutions for this type of problem. In addition, previous studies have shown that clinicians are not aware of the complex interaction among variables that a neural network can exploit. Two separate studies have compared the accuracy of neural network with that of clinicians to predict disease or outcome [30-31]. In the first study, emergency department physicians and medical residents were asked to identify myocardial infarction in patients presenting at an emergency department based on clinical and ECG findings. Eight of 36 cases of myocardial infarction were missed by physicians, compared with only 1 case missed by using the neural network, yielding sensitivities of 77.7%(95%CI, 77 to 82.9%) and 97.2% (95% CI, 97.2 to 97.5%), respectively. In another scenario, the overall accuracy of physicians to predict outcome for colorectal cancer ranged from 75 (95% CI, 66 to 84%) to 79% (95% CI, 71 to 87%), compared to 90% (95% CI, 84 to 96%) for the neural network. The superior prediction capability of neural networks over

physician assessment was observed also in this study, which implies that the complexity of biological systems may be beyond the analytic capabilities of physicians. In our study the sensitivity and the specificity for the 4th year survival status was 92.8% and 81.1% and for the 5th year, it was 97.4% and 78% respectively. H.B.Burke et.al concluded logistic regression and the back propagation neural networks are more accurate in predicting the breast cancer survival. Even our findings are similar to H.B.Burke et.al and Delen Waller and Kadam (2005) for 5 year breast cancer survival rate.

An essential component of the present study is the ability of the neural network to generalize to new population samples. This feature is, however, affected by many factors, such as number of neurons is the hidden units, the type of connections in the network, and the extent to which the network has been trained. The results obtained from the validation set indicate that the network described herein, may generalize well to new patient data. Another advantage of the neural network is its ability to handle missing values. In logistic regression, missing values are usually omitted from further analysis. The ANN model performs better than logistic model and CART model.

Neural networks have the ability to approximate predictive output to any desirable degree of accuracy when provided with enough running time. This could result in over fitting, particularly when there is an attempt to increase the processing power of the network by adding a large number of hidden neurons. In this case, the network will end up learning not only the training set but also the noise in the data, which leads to poor generalization. Until the model is tested on a different population set, the study can be viewed only as the first attempt in the use of connectionist models in the predictions of breast cancer. Our study has several implications regarding the clinical application of artificial neural networks as a diagnostic tool for breast cancer. The use of the neural network could provide physicians and health-care workers with a simple and fast tool with which to assess the risk of breast cancer in any patient presenting at a health-care facility.

References

[1] Wu Y, Giger ML, Doi k, et al: "Artificial neural networks in mammography: application to decision-making process in the diagnosis of breast cancer", Radiology, vol.187, pp.81-87, 1993.

[2] Ortiz J, Sabbatini RM, Ghefter CG, Silva CE, "Use of artificial neural network in survival evaluation in heart failure", Arq Bras Cardiol, vol. 64,pp.87-90, 1995.

[3] Gabor AJ, Seyal M, "Automated interictal EEG spike detection using artificial neural network," Electroencephalograph clin Neurophysiol, vol.83, pp.271-280, 1992.

[4] Fogel DB, Wasson EC 3rd, Boughton EM," Evolving neural networks for detecting breast cancer," Cancer Lett, vol.96, pp.49-53, 1995.

[5] Faraqqi D and Simon R, "A neural network model for survival data," Statistics in medicine, vol.14, pp.73-82, 1995.

[6] Ravdin P.M, and Clark G.M, "A practical application of neural network analysis for predicting outcome of individual breast cancer patients", Breast cancer Research and Treatment, vol. 22, pp.285-293, 1992.

[7] Cruz J.A, Wishart D.S," Application of machine learning in cancer prediction and prognosis", Cancer Informatics, vol.2, pp.59-77, 2006.

[8] Abbass H.A, An evolutionary artificial neural network approach for breast cancer prognosis, Artificial Intelligence in Medicine, vol.25, pp.265-281, 2002.

[9] Lisoba PJ, Etchells TA, Jarman IH, Hane Aung MS, Chabaud S, Bachelot T, Perol D, Gargi T, Bourdes V, Bonnevay S, Negrier S, "Time-to-event analysis with artificial neural networks: an integrated analytical and rule-based study for breast cancer", Neural Netw, vol. 21(2-3), pp.414-426, Mar-Apr 2008.

[10] S.F.Brown, A.J.Branford, W.Moran, "On the use of artificial neural network for the analysis of survival data", IEEE Transactions on Neural Networks, vol.8, pp.1071-77, 1997.

[11] H.B.Burke, et al, "Artificial neural networks improve the accuracy of survival prediction", Cancer, vol.79, pp.857-862, 1997.

[12] Astin ML, wilding P, "Application of neural network to the interpretation of laboratory data in cancer diagnosis", clin chem., vol.38, pp.34-38, 1992.

[13] Venkatesan P. and Anitha S, "Application of a radial basis function neural network for diagnosis of diabetes mellitus". Vol.91, pp.1195-99, 2006.

[14] Venkatesan P. and Suresh M.L. "Neural Network model for classifying renal failure", Mathematics Computing and Modeling, Allied Publishers, India pp.150-160, 2007.

[15] Astion M.L, Wener M.H, Thomas R.G, Hunder G.G and Bloch D.A, "Over training in neural network that interpret clinical data", Clin Chem. vol. 38, pp.1998-2004, 1993.

[16] Wingo PA, Tong T, Bolden S, "Cancer statistics", CA cancer J.clin, vol.41, pp.8-30, 1995.

[17] Imran Kurt, Mevlut Ture, A.Turhn Kurum,"Comparing performances of logistic regression, classification and regression tree and neural networks for predicting coronary artery disease", Expert Systems with Applications, vol34, pp.366-374,2008.

[18] Kenneth R. Hess, Marie C. Abbruzzese, Renato Lenzi, Martin N. Raber, and James L. Abbruzzese, "Classification and Regression Tree Analysis of 100 Consecutive Patients with Unknown Primary Carcinoma", Clinical Cancer Research, vol5,pp.3403-3410,1999.

[19] Lee E.T, Statistical methods for survival data analysis, John Wiley and sons, Newyork, 2003.

[20] Burke, Harry B, Rosen, David B, Goodman, Philip H, "Comparing artificial network to other statistical methods for medical outcome prediction", IEEE International Conference on Neural network Conference Proceedings, vol.4, pp.2213-16, 1994.

[21] Delen D, Walker G, and Dadam A, "Predicting Breast cancer survivability: a comparison of three data mining methods", Journal of Artificial Intelligence in Medicine, vol 34, pp.113-127, 2005.

[22] Bourdes V, Bonnevay S, Lisboa P.J.G, Aung M.S.H, Chabaud S,Bachelot T, Petrol D, Negrier S, "Breast cancer prediction by neural network analysis: A comparison with logistic regression", 29th Annual International Conference of IEEE-EMBS, Engineering in medicine and biology medicine, 4353569,pp.5424-27, EMBC 2007.
[23] L.Breiman,J.Friedman, R.Olshen and C.Stone, Classification and Regression Trees, Wadsworth and Brooks, 1984.
[24] D.E.Rumelhart, G.E.Hinton and R.J.Williams, "Learning representations by back propagating errors", Nature, vol.323, pp.533-536, 1986.
[25] R.Hecht-Nielsen, Neurocomputing, Addision-wesley, Reading, A, 1990.
[26] H.White and A.R.Gallant, "On learning the derivatives of an unknown mapping with multiplayer feed forward networks", Neural Networks, vol.5, pp.129-138, 1992.
[27] H.White, Artificial Neural Networks. Approximation and learning theory, Blackwell, Cambridge, MA, 1992.
[28] C.M.Bishop, Neural Networks for pattern recognition, Oxford University Press, New York, 1995.
[29] Flexer A. "Statistical evaluation of neural networks experiments: Minimum requirements and current practice", In Proceedings of the 13th European Meeting on Cybernetics and Systems Research, Vienna, Austria vol. 2,pp.1005-08, 1996.
[30] Baxt WG. "Use of an artificial neural network for the diagnosis of myocardial infarction", Ann Intern Med, vol.115, pp.843-848, 1991.
[31] Bottaci L, Drew PJ, Hartley JE, et al. "Artificial neural network applied to outcome prediction for colorectal cancer patients in separate institution", Lancet ,vol.350,pp.469-472, 1997.

received the M.Sc.,M.Phil. and PhD degrees, from Madras Univ. in 1975, 1980 and 1990 respectively. After working as Asst.statistician (from 1982), Research Officer (from 1991), Senior Research Officer (from 1995), Assistant Director (from 2001), and Deputy Director (from 2006) in the Department of Statistics, Tuberculosis Research Centre (TRC), Indian Council of Medical Research (ICMR), Chennai. He is now working as a Scientist at TRC, ICMR since 2008. His research interest includes machine learning, computational biology, survival analysis, disease modeling, hidden Markov models and MCMC. He is the fellow of ISMS, India. He is a member of IBS, IABMS, ISPS, and IASP

received the M.Sc., and M.Phil from Madurai Kamaraj University in 1984 and 1990 respectively. After working as Lecturer (from 1984), Assistant professor (from 1996) in Sathyabama University, in the Department of Mathematics, he is now working in Easwari Engineering College, as Assistant professor and doing his research activities under the guidance of Dr.Venkatesan. His areas of interest are neural network and fuzzy logic.