Prognostic Systems for NPC: A comparison of the Neural Network Model and The Cox Proportional Hazards Model

Sameem Abdul-Kareem[1], Sapiyan Baba[2], Yong Zulina Zubairi[3], U Prasad[4], Mohd Ibrahim A Wahid[5]

[1] Faculty of Computer Science & Information Technology
University of Malaya, Kuala Lumpur
[2] Centre for Information Technology, University of Malaya
[3]Institute of Mathematical Sciences, University of Malaya
[4]ENT Department, Faculty of Medicine, University of Malaya Medical Centre,
[5]Pantai Medical Center, Kuala Lumpur
MALAYSIA

Abstract

Statistical methods such as the life-table, the Kaplan-Meier method and regression models, such as the Cox Proportional Hazard are usually used to model and predict survival data. Neural networks have been used in medicine for more than two decades, first, as an aid to diagnosis and treatment and then, recently, as a tool to study medical prognosis of aids, coronary heart disease, and a variety of cancer types. The use of neural networks to study the prognosis of nasopharyngeal carcinoma is, however, relatively new. In this paper we describe our research in the use of neural network to predict the prognosis of nasopharyngeal carcinoma. Two prognostic models for nasopharyngeal carcinoma were developed, namely the neural network model and the Cox model and their performance compared.

Keywords: artificial neural network, back propagation, nasopharyngeal carcinoma, survival analysis, prognosis, Cox proportional hazards

1 Introduction

Cancer of the nasopharynx also known as nasopharyngeal carcinoma (NPC) is an abnormal malignant growth of the tissue and the mucosa lining the nasopharynx. [1, 2, 3, 4]. The nasopharynx is an area behind the nose and the upper part of the throat above the soft palate at the back of the mouth. NPC is the most common cancer of the head and neck in the southeastern part of China, Taiwan, Hong Kong, Malaysia and Singapore. In Malaysia, NPC is also the most common cancer.

Nasopharyngeal carcinoma is a tumour that is predominant amongst the Chinese and South East Asian ethnic community. The incidence of NPC amongst the Chinese community of Malaysia is approximately 27 per 100,000 compared to China at 40 per 100,000 and Hong Kong at 35 per 100,000 population. In the United States and Europe the incidence is only about 1 per 100,000 [5, 6, 7].

To date, there is very little published work (if any) on the prognosis of NPC. Medical prognosis is a prediction of the future course and outcome of a disease and an indication of the likelihood of recovery from that disease. However, it is only a prediction and like all predictions it is not 100% accurate.

Cancer prognosis is difficult because different patients cannot be observed for the same length of time, known in prognosis terminology as censorship. Statistical methods such as the life-table, the Kaplan-Meier method and regression models such as the Cox Proportional Hazard are usually used to model and predict survival data with the ability to handle censored data.

More recently, experiments have been performed on using alternative methods for the analysis of survival data on a variety of diseases. One such method is the use of artificial neural networks (ANNs).

The analysis is based on NPC cases seen in the UMMC, Kuala Lumpur from 1969 to 1999. This collection of data was transcribed from paper to an electronic media in 1982. The design of the database was based on a similar database used in Mayo Clinic. At present the dataset has a total of 1693 cases. Variables include age, sex, race, dialect, date first seen, type seen, biopsy, diagnosis, symptom, tumour extent, nerve involvement, node involvement, distant metastasis, WHO Type, TNM classification and stage.

2 Data Analysis

Data from 514 patients were made available for the purpose of this research. The data shows that NPC is more common in male than in female with a ratio of 2:1 [8].

The disease is common amongst the Chinese ethnic group at 88% as compared to the Malay ethnicity of only 10.5% and the Indian ethnic group of merely 1.5%. The incidence of the NPC amongst the 3 ethnic groups is in agreement with the findings reported in [5,6,7]. The following statistic is a summary of the dataset in the study:

Table 1 Data Analysis by Stage

Stage I	Stage II	Stage III	Stage IV
5%	2.5%	2.5%	90%

Table 2 Data Analysis by WHO Type

Type I	Type II	Type III
11.9%	15.9%	72.2%

Table 3 Data Analysis by Tumour Type

T1	T2	T3	T4
35%	14%	10%	41%

Table 4 Data Analysis by Node Size

N0	N1	N2	N3
15.2%	12.8%	50.8%	21.2%

Table 5 Data Analysis by M

M0	M1
89%	11%

Table 6 Data Analysis by Treatment Type

Adjuvant Therapy	Neoadjuvant Therapy	Radiotherapy	Chemotherapy
3.5%	8.5%	87%	1%

3 The Cox Regression Model

The Cox Regression (Cox Proportional Hazards) model is the most general of the regression models in the sense that it does not make any assumption about the nature or the shape of the underlying function. The model assumes that the underlying hazard rate (instead of the survival time) is a function of independent variables [9]. The function can be expressed as

$$h(t) = h_0(t) \exp (b_1 z_1 + b_2 z_2 + \ldots b_m z_m), \quad h_0(t)\exp (\Sigma_i^m b_i z_I) \qquad (1)$$

where $h(t)$ is the hazard function of the respective survival time (t), the z_i's are the prognostic factors and h_0 is the *baseline hazard*, that is the hazard for the respective individual when all independent variables (z variables) are equal to zero, that is, the hazard function of the underlying survival distribution when all the z variables are ignored.

The Cox Regression procedure from SPSS 10 was used to build the Cox model. Variable selection was done using the Wald backward selection. Estimation of the survival at event times was performed using the SPSS SAVE Survival Function procedure.

Using a threshold of 0.5 a patient is considered dead at a given interval if the probability survival falls below this value. This was compared against the survival information given in the actual database.

4 The Neural Network Model

Survival analysis using the Cox proportional hazards generally looks at the population as a whole and hence cannot prognosticate at individual level. In the clinical scenario, this is what the patient expects. Most patients would like to have a more accurate prediction as a result of therapy. They want to know how well they will fare on the various therapies. Recently, experiments have been performed on using alternative methods of artificial neural network technology for the analysis of survival data on a variety of diseases [10, 11, 12, 13, 14, 15]. The main advantage of neural network technology is that the internal representation and distribution of data need not be known. Although neural networks has not been tested extensively for modelling survival data, based on its predictive successes in other domains, it is considered a good alternative for the prediction of survival of individual patients. Neural networks also do not offer any obstacle to handling censored data [13, 14].

4.1 Pre-processing

To make the neural network training more efficient a pre-processing is performed on the network inputs and targets. One of the pre-processing techniques recommended is a procedure that scales the network inputs and targets by normalising the mean and standard deviation of the training set. The inputs and targets are normalised so that they have zero mean and unity standard deviation.

In some situation, the dimension of the input vector is large but the components of the vectors are highly correlated and are therefore redundant. Principle component analysis (PCA) is a procedure used to reduce the dimension of the input vectors. PCA orthogonalises the components of the input vectors so that they are uncorrelated; it orders the resulting orthogonal components (principal components) so that those with the largest component comes first; it eliminates those components which contribute the least to the variation in the data set.

4.2 Post-processing

The performance of both the Cox model and the neural network was measured in terms of the area under the receiver-operating characteristic and by the number of correctly predicted survival. In medical prognosis prediction ROC is commonly used to determine the accuracy of predicted values as it can be used across different classification tools.

The ROC is a plot of sensitivity versus specificity for different test results. A person with the disease who has a "positive" test result is termed a true positive, whereas a person with the disease but a "negative" test is termed a false negative. On the other hand, a person without the disease who has a "positive" result is termed a false positive, while person without the disease but a "negative" test is termed a true negative [16].

In this case, true positive can be defined as the person who is dead and has a positive test result, false negative is the person who is dead and has a negative result. False positive is a person who is alive but has a positive result and true negative is a person who is alive and has a negative test result. Table 7 summarises this description.

Table 7 The Definition of True Positives/Negatives

	Dead (+)	Alive (-)
Result Positive	A = true positive	B = false positive
Result Negative	C = false negative	D = true negative

Sensitivity is the true positive test results divided by all the living patients. This is the probability that a patient will be classified as alive when she is alive.

$$\text{Sensitivity} = (a / (a + c)) \qquad (2)$$

The specificity of a test is the true-negative test results divided by all the dead patients. This is the probability that a patient will be classified as dead when she is dead. "1-specificity is the probability that a patient will be classified as alive when she dead.

$$\text{Specificity} = (d / (b + d)) \qquad (3)$$

To generate the ROC curve it is first necessary to determine the sensitivity and specificity for each test result. The X-axis ranges from 0 to 1, or 0% to 100% and is the false positive rate, i.e. 1-specificity. The Y-axis ranges from 0 to 1, or 0% to 100% and is the true positive rate, i.e. the sensitivity. The curve starts at (0,0) and increases towards (1,1). The endpoints of the curve will run to these points and an area of the resulting trapezoids can therefore be calculated. The larger the area under the curve the better is the prediction.

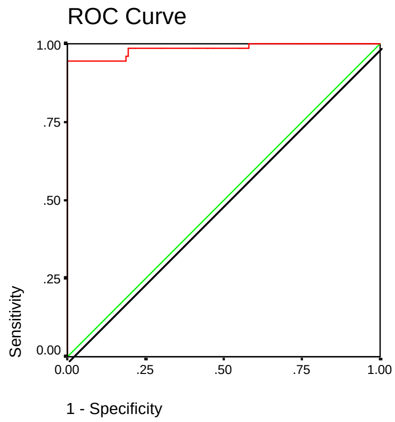

Figure 1 ROC Curve of Year 10 of the Neural Model, ROC Area = 0.957SE = 0.028

The accuracy of predictions is also measured by the number of correctly predicted cases divided by all the cases in the study as summarised by Table 8. Neural networks usually produce real numbers as their outputs. These real numbers are then processed to provide a classification: a threshold is usually established above which the final output of a certain unit will be "1" or true. The percent accuracy of prediction depends largely on the threshold that is used to convert the outcomes into '1' and '0' and also the size of the dataset.

We have also attempted to use a 5-fold cross validation in order to prevent over-prediction.

Table 8 The Accuracy of The Prediction is 5/8 or 62.5%

Actual	Prediction	Outcome
0	0.0001	0
0	0.0212	0
0	0.0411	0
0	0.999	1
1	0.0012	0
0	0.011	0
0	0.965	1
0	0.0081	0

Status according to the standard, "1"- Dead and "0"- Alive

5 Results

Table 9 Comparisons of Neural Network and Cox Proportional Model

Yr	NN			Cox		
	ROC	SE	%	ROC	SE	%
1	0.795	0.062	79.71	0.809	0.021	80.2
2	0.782	0.053	71.64	0.743	0.024	68.9
3	0.838	0.053	81.2	0.726	0.026	63.6
4	0.753	0.057	76	0.703	0.028	67.8
5	0.795	0.051	79	0.689	0.031	56.9
6	0.864	0.045	79.5	0.688	0.037	51.6
7	0.845	0.046	84	0.690	0.038	51.9
8	0.918	0.034	90	0.667	0.046	49.9
9	0. 918	0.036	90	0.664	0.053	49.6
10	0.957	0.028	93	0.704	0.051	50.7

Table 9 shows the results of using a neural network model to predict outcome in NPC and that of using a Cox Proportional Hazards model. Year 1 and Year 2 predictions for both models were almost equal with 0.795 (Year 1) and 0.782 (Year 2) for the neural model and 0.809 (Year 1) and 0.743 (Year 2) with 0.795 (Year 1) and 0.782 (Year 2) for the neural model and 0.809 (Year 1) and 0.743 (Year 2) for the Cox model. However, Year 3 prediction for the neural model showed greater accuracy than that of the Cox model. From Year 3 onwards the neural model predictions showed only improved accuracy while that of the Cox model deteriorated.

Table 10 The difference in resolution between the neural network model and the Cox proportional hazard model.

Year	dnnc
1	-0.014
2	0.039
3	0.112
4	0.05
5	0.106
6	0.176
7	0.155
8	0.251
9	0.25
10	0.253

Resolution measures how much the model is able to separate cases with “true” (positive) outcome from those with “false” (negative) outcome. The area under the receiver operating characteristics curve (AUC) is a graphical representation of resolution.

Table 10 refers to the difference in resolution obtained from the neural model and the Cox model (dnnc). Neural network models provided larger areas under the ROC curve except for the 1st interval, i.e. Year 1. A chi-square test, showed that the differences between the results obtained using the neural method and that using the Cox method, are significant with p = 0.014 ($p < 0.05$).

In general, the results obtained from using ANN increases in accuracy for the predictions of Year 1 to Year 10. The same trend cannot be observed in the results obtained using the Cox Proportional Hazards method (refer to Table 9).

Thus, the results obtained by using neural network is more consistent than that obtained from using the Cox method. Neural networks had better predictive performance than the Cox proportional hazards model in the NPC dataset.

Table 11Adjusted Areas Under the ROC

Year	Adjusted ROC	Unadjusted ROC
1	0.795	0.786
3	0.838	0.820
5	0.795	0.790
7	0.845	0.867
9	0.918	0.903

Table 11 shows the results obtained after having used a cross-validation method. However, a chi-square test shows that there is no statistical difference between (p = 0.24) the unadjusted resolution and that of the adjusted resolution.

6. Conclusion and Future Work

We have shown that the resolution of the Cox model is inferior to that of the neural model (see Tables 9 and 10).

Furthermore, using the Cox model only a retrospective analysis can be done, as the dependent variable, i.e., survival time is required for the analysis to be carried out. We admit that it may be possible, via some extrapolation exercises, to carry out the analysis of new cases using the Cox method but this may be tedious and the accuracy is questionable, as even a direct analysis using the Cox model are less accurate than that of the neural model.

Statistical models are usually used for large groups of people based on estimates taken from a sample and are therefore meaningless for an individual. Although there are attempts in statistical tools such *SPSS* to provide predictions for individual cases, these predictions far from accurate and are affected by the number of similar cases in the sample. As an example if there is only a small number, say, five cases, with a survival of eight years, the prediction is different if there are, say, fifty such cases.

Hence, although a comparison between the neural network model and that of the Cox proportional hazard model was attempted, the comparison was, at the best, a biased one.

On the other hand, through the generalization capability of the neural network, it is possible to train the network based on one set of data and to test the network on another set of data without having to input the dependent variable, which in this case, is the survival time. Thus, neural networks are able to predict new cases of NPC, in which the survival time is as yet unknown. Using neural networks we are also able to make predictions for individual cases by looking at the results of a separate test set

From the results in Tables 9 and 10 we have shown that neural network models produces better estimates of survival than the Cox proportional hazards model.

Other work that can be carried out in future would be to incorporate censored data in a similar research. There have been several approaches to this problem by different researchers, as reported in [15, 17, 18, 19].

As reported by B.D. Ripley and R.M. Ripley in [20], when censored cases are omitted from the network, predictions of survival will be biased downwards as the censored patients are considered as dead although they may still be alive. Thus, we could expect an improvement in the results obtained in this research if censored cases are considered in the training of the network.

Reference

[1] DeNittis, A., Dilling, T. and Goldwein, J., “The biology of cancer, 1999. http://oncolink.upenn.edu/specialty/med_onc/nature_cancer1.html

[2] NCI PDQ “What is cancer of the nasopharynx?, Nasopharyngeal Cancer, 1999. http://cancernet.nci.nih.gov/clinpdq/pif/Nasopharyngeal_cancer_Patient.html#5

[3] American Joint Committee on Cancer (AJCC), *AJCC Cancer Staging Manual*, ed. I.D. Fleming, J.S. Cooper, D.E. Henson, R.V.P. Hutter, Pennsylvania: Lippincott-Raven.1998, pp. 1-41.

[4] Dunn, S., 1997. *Basic Information On Cancer*, http://cancerguide.org/basic.html

[5] Chia, K.S. and Lee, H.P. 1997,“Epidemiology” in Nasopharyngeal Carcinoma, 1997, eds V.H. Chong, and S.Y. Tsao, pp.1-5. Singapore: Armour Publishing.

[6] Prasad, U. & Rampal, L., “Descriptive Epedimiology of Nasopharyngeal Carcinoma in Peninsula Malaysia”, Cancer Causes Control, vol. 3, 1992, pp. 179-182.

[7] Shanmugaratnam, K. & Tye, C., “A study of nasopharyngeal cancer among Singapore Chinese with special reference to migrant status and specific community

(dialect group)". Journal of Chronic Diseases vol. 23, 1970, pp. 433-441.
[8] Perez, A., "Nasopharynx" in Principles and Practice of Radiation Oncology, 2nd Ed., eds A. Perez & L.W. Brady, Philadelphia: J.B. Lippincott Co, 1992, p. 617-643.
[9] Cox, D.R. & Oakes, D., Analysis of Survival Data, Cambridge, Chapman and Hall, 1994.
[10] Burke, H.B., Rosen, D.B, and Goodman, P.H., "Comparing the prediction accuracy of statistical models and artificial neural network in breast cancer". Proceedings of the NIPS 94 Post-Conference Workshop, 1994.
[11] Burke, H.B., Rosen, D.B, and Goodman, P.H., "Artificial neural networks improve the accuracy of cancer survival prediction". Cancer vol. 79, 1997, pp.857-62.
[12] Burke, H.B., Rosen, D.B, and Goodman, P.H., "Predicting response to adjuvant and radiation therapy in patients with early stage breast carcinoma". Cancer vol. 82, 1998, pp. 874-877
[13] Ohno-Machado, L., "Medical applications of neural networks: connectionist models of survival". PhD Thesis, 1996, Section in Medical Informatics, Stanford University. Palo Alto, California.
[14] Ohno-Machado, L., "Neural network techniques: utilization in medical prognosis", Electronic Copy of Review, 1999. [Emailed by Ohno-Machado]
[15] Street, W. N., "A neural network model for prognostic prediction", 1999. http://dollar.biz.uiowa.edu/~street
[16] Evans, C., 1999. http://www.psyctc.org/stats/roc1/
[17] Lapuerta P, Azen S., LaBree L., "Use of neural networks in predicting the risk of coronary-artery disease". Computers and Biomedical Research, vol. 28, 1995, pp.38–52.
[18] Ravdin, P.M. & 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, no. 3, 1992, pp.285-293.
[19] Ravdin .P.M, Clark G.M., Hilsenbeck S.G., Owens, M.A., Vendely P., Pandian, M.R., McGuire, W.L., "A demonstration that breast cancer recurrence can be predicted by neural network analysis". Breast Cancer Research and Treatment, vol. 21, 1992, pp. 47–53.
[20] Ripley, B.D., & Ripley, R.M., "Neural networks as statistical methods in survival analysis", http://www.stats.ox.ac.uk/~ripley/papers.html