ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

A comparative study of artificial neural network architectures and cox proportional hazard model using heart attack data

S. Prathap[1*], V. Vallinayagam,[2] and P. Venkatesan[3]
[1]Department of Mathematics, Jeppiaar Engineering College, Chennai, TamilNadu, India.
[2]Department of Mathematics, St. Joseph's College of Engineering, Chennai, TamilNadu, India.
[3]Department of Statistics, Sri Ramachandra University, Chennai, TamilNadu, India.
***Corresponding Author: prathapmaths@gmail.com**

ABSTRACT

An interrelated group of artificial neurons that uses a mathematical model for information processing based on a connectionist approach to calculation is called an Artificial Neural Network (ANN). Performance of the Radial Basis Function Neural Network (RBFNN) was also compared with the most commonly used Multi-Layer Perceptron (MLP) network model and the Cox Proportional Hazard (PH) model. Heart attack database was used for empirical comparisons and the outcomes show that RBFNN performs better than other models.

KEYWORDS: Cox proportional hazard, artificial neural network, multilayer preceptor, radial basis function, supervised learning

1. INTRODUCTION

The Cox PH model is broadly used for the study of time to event data in the presence of covariates. This Cox model is popular because of its simplicity, and not being based on any assumption about the survival distribution. The theoretical basis for the model has been solidified by linking it to the study of counting processes and martingale theory, given in the books of Fleming and Harrington and Andersen. These growth have led to the introduction of numerous new extensions of the original model.

Neural computing is an information processing concept, motivated by biological system, framed of a large number of highly interrelated processing elements or neurons to solve particular problems. An ANN is configured for a particular application, such as pattern recognition or classification of data, done a learning process. Learning in biological systems requires adjustments to the synaptic connections that exist among the neurons. The McCulloh and Pitts network had a fixed set of weights. In 1949 Hebb modernized the first learning rule, i.e., if two neurons are participating at the same time then the lastingness between them should be increased. In 1950 and 1960's Block, Minsky, Papert, and Rosenblatt are exploited on perceptron. The Neural Network (NN) model could be proved to converge to the correct weights that will resolve the problem. By Herb, the weight adjustment or learning algorithm used in the perceptron was establish more potent than the learning rules. Parker and Lacuna discovered a learning algorithm for MLP networks called Back Propagation (BP) that could solve problems for non-linearly separable.

In Heart attack works, the major outcome of interest is the time to occurrence of event like death, relapse etc. Heart failure occurs when the pumping action of the heart cannot give enough blood to the part of the body as it is required. Heart works tough to pump blood around the body and causes blood to flow at increased pressure through the blood vessels, causing risk for a heart attack or stroke. High Blood Pressure is a main hazard factor for coronary heart disease and stroke by Gibbs *et al*., 2000. Stroke occurs while a blood vessel supplying blood to one branch of the brain becomes blocked. As a result, no oxygen can get to this branch of the brain and leads to death. Symptoms of stroke incorporate failing in the arm, leg, hand or face, sightlessness, trouble in talking and loss of balance by Schoenstad, 2008. In 2007, American Heart Association reported that cardiovascular disease is the single largest reason of death among women in the America and worldwide, accounting for one-third of all deaths. In fact, more women than men die every year of cardiovascular disease.

The aim of this study is to compare the performance of MLP, RBFNN and Cox PH model using Heart attack data.

2. MATERIALS AND METHODS

Cox regression analysis: If a set of covariates is represented by $Z_i = (z_{1i,} z_{2i}, \dots, z_{ki})$, famous Cox regression model is

$$h_i(t, z_i) = h_0(t) \exp\left(\sum_{j=1}^{k} \beta_j z_{ij}\right), \qquad j = 1, 2, \dots, k \tag{1}$$

where $h_0(t)$ is the hazard function of time only, assume to be similar for all subjects, β_j are unknown parameters, describe the significance of covariates and z_{ij} are the values measured on subject i at time zero. The baseline hazard describes the form of the distribution while $\exp(\beta z_i)$ gives the level of each individual's hazard. The Cox model based on the assumption that independent covariates effect the hazard in a multiplicative way. z is a vector of covariates of interest. z may be discrete, continuous or the mixture of both. The main advantage of Cox PH model is that we can estimate the parameterβ, without having to calculate $h_0(t)$. This fact makes the model as a semi-parametric. Cox regression analysis, was used as a comparison for the performance of NN approach. The relationships between different covariates and patient survival with the calculation of the forecast of the patient outcome, were assessed by applying Cox PH regression.

Artificial Neural Network (ANN): ANN at first developed to imitate basic biological neural systems, the human brain particularly, are collected of a numeral of interlinked simple processing elements called nodes or neurons. Every node gets an input sign which is the total information from other nodes, processes it locally though an activation or transfer function and produces a transformed output sign to other nodes. Every single neuron enforces its function rather slowly and badly, jointly a network can execute an amazing number of jobs efficiently Reilly and Cooper. This information processing characteristic create ANN a powerful mathematical device and able to find out from examples. First studies of NN were done in 1942 by McCullough and Pitts. After sometime, Rosenblatt conceived in 1959 the first learning algorithm, creating a model known as the perceptron, which was then only a solution to simple linear problems. Werbos reported, the first non-linear processing capabilities of ANNs in 1974.

Multilayer preceptor (MLP): In MLP, the weighted sum of the inputs and bias terms 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 Forward Neural Network (FFNN). The diagrammatic representation of FFNN is given in Fig.1. The three layers of ANN are input layer, hidden layer and output layer. The learining power of the MLP, increases by the hidden layer. The activation function of the network changes the input to provide a desired output. The activation function is choses by the algorithm require a function with a continuous, single-valued with first derivative existence. Choice of the numeral of hidden layers, hidden nodes and type of activation function acting an essential position in model building, Hecht-Nielsen, and White.

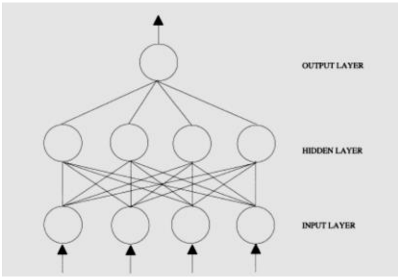

Fig.1 Feed forward neural network

Training of Multilayer preceptor neural networks: Back Propagation (BP) algorithm is the generalization of the least square algorithm. MLP is a supervised learning and it occurs in the preceptron by varying connection weights. The changes of each weight $\Delta W_{ji}(n)$ can be computed by gradient descent as in equation (2)

$$\Delta W_{ji}(n) = \eta \frac{\partial z(n)}{\partial v_j(n)} y_i(n) \tag{2}$$

where $y_i$is the output of the previous neuron and η is the learning rate. This derivative can be simplified as

$$-\frac{\partial z(n)}{\partial v_j(n)} = e_j(n)\,\phi'\left(v_j(n)\right) \tag{3}$$

$$-\frac{\partial z(n)}{\partial v_j(n)} = e_j(n)\,\phi'\left(v_j(n)\right)\sum k - \frac{\partial z(n)}{\partial v_j(n)}\, W_{kj}(n) \tag{4}$$

where ϕ' is the derivative of the activation function and the error can be find out by

$$e_j(n) = d_j(n) - y_i(n) \tag{5}$$

where y is the value as result of perceptron and d is the target value. The correction to the weights based on those correction which minimize the error in the entire output $\varepsilon(n)$ is given by equation (6) below

$$\varepsilon(n) = \frac{1}{2}\sum e_j{}^2(n) \tag{6}$$

This error signal is used to revised the weights and thresholds of the hidden layers and the output layer. The weights and the thresholds are revised in an iterative method until the error signal becomes minimum and the Mean Square Error (MSE) is taken as a performance measurement.

Radial Basis Function (RBF): The RBF was first proposed by Powell (1985), to resolve the real multivariate interpolation problem. But RBF was first used by Broomhead and Lowe. Park and Sanndberg, proposed the approximation theorem, which played a vital role in practical application of RBF. RBFNN based on supervised learning. An alternative to MLP is RBF. RBF networks are also good at modelling non-linear data and can be trained in one stage rather than using an iterative process as in MLP. RBF are useful in solving problems where the input data are spoiled with additive noise. Based on Gaussian distribution, the transformation functions are used.

RBF Network Model: The RBF network has a feed forward composition lying of a one hidden layer of J locally tuned units, which are completely interconnected to an output layer of L linear units. All hidden units at the like time receive the n-dimensional real valued input vector X the major difference form that of MLP is the lack of weights of hidden-layer. The hidden-unit outputs are not estimated using the weighted sum mechanism /sigmoid activation, rather each hidden unit output Z_j is obtained by closeness of the input X to an n-dimensional parameter vector μ_j associated with the j^{th} hidden unit. The response characteristics of the j^{th} hidden unit $(j = 1, 2,3, \ldots, J)$ is assumed as,

$$Z_j = K\left(\frac{\|X - \mu_j\|}{\sigma_j^{\,2}}\right) \tag{7}$$

where K is symmetric function with maximum at its centre μ_j. The parameter σ_j is the breadth of the respective field in the input space from unit j. Z_j has an considerable value only when the distance $\|x - \mu\|$ is the lesser than the breadth σ_j.

Given the input vector X, the output of the RBF network is the L-dimensional activity vector Y, whose l^{th} component $(l = 1, 2, 3, \ldots, L)$ is given by

$$Y_l(X) = \sum_{j=1}^{J} W_{lj}\, Z_j(X) \tag{8}$$

From equation (1) and (2) , we can observe that, the RBF produced network structure, which is similar to a desired function $f(X)$ by superposition of non-orthogonal, bell formed basis functions. The overall accuracy of RBF networks can be controlled by three parameters, the numeral of basis functions, their location and their width . Assumed a Gaussian basis function for the hidden units given as Z_j for $j = 1, 2,3, \ldots, J$.

$$Z_j = exp\left(-\frac{\|X - \mu_j\|^2}{2\sigma_j^{\,2}}\right) \tag{9}$$

μ_j and σ_j are mean and the standard deviation respectively.

Training of RBFNN: Training of the RBFNN is a two-step learning strategy. First, the centers of each of the J Gaussian basis functions were fixed to characterize the density function of the input space. The second is to find out the weight vector W which would best approximate the limited sample data, thus leading to a linear optimization problem that could be solved by ordinary Least Square (LS) method. LS function is the common objective function which helps to select the parameter values that minimize its values. The minimization of the LS objective function by optimal choice of weights optimizes accuracy of fit. If we have several objective like smoothness and accuracy a regularized objective function is preferable. The sum of squared error (SSE) criterion function can be considered as an error function E by adaptively updating the free parameters of the RBF network. Three types of RBF network that require to be chosen to adapt the network for a particular job. They are receptive field centers μ_j of the hidden layer, the receptive field breadths σ_j and the output weights W_{ij}. If we use fully supervised gradient-descent method training, then μ_j, σ_j and W_{ij} are updated as follows.

$$\Delta\mu_j = -\rho_\mu\, \nabla_{\mu j} E \tag{10}$$

$$\Delta\sigma_j = -\rho_\sigma\, \frac{\partial E}{\partial \sigma_j} \tag{11}$$

$$\Delta W_{ij} = -\rho_W\, \frac{\partial W}{\partial W_{ij}} \tag{12}$$

3. APPLICATION TO HEART ATTACK DATA

Database: The data obtained from the website, ftp//ftp.wiley.com/public/scitech_med/survival and http://www.umass.edu/statdata/statdata. The data from the Worcester Heart Attack Study (WHAS) have been provided by Goldberg (1989) of the Department of Cardiology at the University of Massachusetts Medical School. Data collected from 1975 to 1988, on all myocardial infarction (MI) patients allowed to hospital in the Worcester, Massachusetts Standard Metropolitan Statistical Area. Event is encrypted as 1 and censoring is encrypted as 0. The subsets of covariates used, with their codes and values for provided in Table.1.

For MLP network architecture, one hidden layer with activation function of sigmoid, which is optimal for the outcome, is chosen. A BP algorithm grounded on conjugate gradient optimization technique was used to model MLP for the above data. A Cox PH model was fitted using the same input vectors as in the neural networks and heart attack status as the binary dependent variable. Constructed models efficiency was evaluated by likening the sensitivity, specificity and overall accurate predictions for datasets. Cox PH, MLP and RBFNN were constructed using SPSS and MATLAB Softwares.

Table 1. Description of the covariates obtained from WHAS

Covariates	Description	Codes/units
AGE	Age	years
SEX	Gender	0 = Male, 1 = Female
CPK	Peak Cardiac Enzyme	Int. units
SHO	Cardiogenic shock complications	0 = No, 1 = Yes
CHF	Left Heart Failure Complications	0 = No, 1 = Yes
MIORD	MI Order	0 = First, 1 = Recurrent
YRGRP	Grouped Cohort Year	1 = 1975 & 1978 2 = 1981 & 1984 3 = 1986 & 1988

4. RESULTS

WHAS data set with 481 records were used for the study. The Cox PH model were fitted using SPSS. The covariates AGE, SHO and CHF are significantly connected with the time to event. The Cox PH regression fitted to the data gave a sensitivity of 85%, specificity 82% and overall accurate predicition of 83%. The MLP architecture had six input variable and one hidden layer with three hidden nodes and one output node. The best MLP was obtained at lowest Root Mean Square (RMS) of 0.2125. MLP sensitivity was 92%, specificity was 91% and percentage accurate prediction was 91%. RBFNN executed best at ten centres and maximum number of centres tried was 18. Root Mean Square Error (RMSE) using the best centres was 0.3212. Sensitivity of the RBFNN model was 97%, specificity was 97% and the percentage accurate prediction was 97%. Execution time of RBFNN is lesser than MLP and when compared with Cox PH model.

Table.2.Comparative predictions of three models

Model	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**
Cox PH	85	82	83
MLP	92	91	91
RBFNN	**97**	**97**	**97**

5. DISCUSSION

MLP and RBFNN are the comprehensively used FFNNs. Both differ basically in the way how the hidden units combine values approcahing from the inputs. The MLP use inner products and the RBF use Euclidian distance. We used both RBF and MLP algorithms for the forecast of heart attack data.

Venkatesan P and Suresh M. L study gives show that the accuracy of ANN for the breast cancer survival forecast was better than regression models, Burke H. B et al. study concluded that neural networks are more accurate in predicting the breast cancer than LR and CART models for fifth year survival. Hacib T in his paper states that RBFNN identifies the parameters of electromagnetic, faster than MLP neural network. Dan Ardelean et al. in their paper reported that quality of the RBF model is better than the quality of the MLP. Padmavathis paper of breast cancer forecast used RBF and MLP, suggested that RBF have good predictive capabilities and time taken was less when compared to MLP. Venkatesan and Anitha study gives that performance of the RBFNN has a better performance than other models like MLP and classical logistic regression. Sereno F et al. paper entitled, comparative study of RBF and MLP neural nets in the Estimation of the Foetal weight and length, concluded with the slight confusion regarding prediction performance while comparing the RBF and MLP networks to resolve the problem of foetal weight prediction. Many researchers have compared the efficiency of RBF and MLP and majority have recommended that RBF network was better than MLP, and some of them doubt the prediction efficiency.

6. CONCLUSION

The sensitivity and specificity of both NN models had a better predictive power compared to Cox PH regression. Also the time taken by RBF is less than that of MLP in our findings. The limitation of the RBFNN is that it is more sensitive to dimensionality and has larger difficulties if the numeral of units is huge. The forecasting capabilities of RBFNN has showed better outcomes and more applications would bring out the efficiency of this model over other models.

REFERENCES

Andersen P K, Borgan O, Gaill RD and Keiding N, Statistical methods based on Counting Processes, Springer, New York, 1993.

Broomhead D, S. And Lowe D, Multivariate functional interpolation and adaptive networks. Complex Syst., 2, 1988, 321-355.

Burke H B, Goodman P H and Rosen D B, Comparing the prediction accuracy of artificial neural networks and other statistical model for breast cancer survival Advances in Neural information processing system, 7, 1995, 1063-1067.

Dan Ardelean, Marius Kloetzer and Octvian Pastravanu, RBF neural networks in nonlinear system identification, Proc 7th International symposium on Automatic Control and Computer Science, 2001, 82-86.

Fleming TR and Harrington DP, Counting Processes and Survival Analysis, John Willey & Sons, New York, 1991.

Hacib T, Mekideche, M R and Ferkha N, Inverse problem methodology for the measurement of the electromagnetic parameters using MLP neural network, word academy of science engineering and technology, 38, 2008, 608-613.

Hartman EJ, Keeler JD And Kowalski JM, Layered neural networks with Gaussian hidden units as universal approximators. Neural Comput., 2, 1990, 210-215.

Haykin S, Neural Networks, A Comprehensive Foundation (2nd ed.) Prentice Hall, 1998.

Hecht-Nielsen R, Neuro Computing. Addisson-Wesley, Reading MA, 1990.

Marubini and Valsecchi, Analysing Survival Data from Clinical Trials and Observational Studies, John Wiley & Sons, 1995.

Padmavathi J, A comparative study on breast cancer prediction using RBF and MLP, International Journal of Scientific and Engineering Research, 2(1), 2001, 1-5.

Parl J And Sandberg IW, Universal approximation using radial basis function networks. Neural Comput., 3, 1991, 246-257.

Poggio T And Girosi F, Networks for approximation and learning. Proc. IEEE, 1990, 78, 1481-1497.

Ramuhalli P, Udpa L and Udpa SS, Finite element neural networks for solving differential equations, IEEE Trans. Neural Networks, 16 (6), 2005, 1381-1392.

Reilly DL and Copper LN, An Overview of neural network: Early models to real world system, in An Introduction to neural and Electronic Networks, Academic Press, San Diego, 1990, 227-24.

Sereno F Marques, De SaMatos. Bernardes, A Comparative Study of MLP and RBF Neural Nets in the Estimation of the Foetal Weight and Length Proceedings of RECPAD. 11th Portuguese Conference on Pattern Recognition, 2000.

Vallinayagam V, Prathap S, and Venkatesan P, Non-Linear Regression Models for Heart attack Data- An Empirical Comparison, Indian Journal of Applied Research, 4(6), 2014, 332-334.

Venkatesan P and Anitha S, Application of Radial basis function neural network for diagnosis of diabetes mellitus, Current Sceience, 91(9), 2006, 1195-1199.

Venkatesan P and Suresh ML, Brest Cancer survival prediction using Artificial Neural Network, IJCSNS International Journal of Computer Science and Network Security, 9 (5), 2009, 169-174.

Venkatesan P and Tintu Thomas, A Comparitive Study of Neural network architecutres for predicting Gene Expression in M. Tuberculosis, International Journal of Pharma and Bio Sciences, 4(1), 2013, 965-972.

White H, Artificial Neural Networks. Approximation and Learning Theory, Blackwell, Cambridge, MA, 1992.